

Db 233 TTGGAATGGCCGAGNCTTGCTAATTCCTGCTTCCACGATCATTAACATTAC 292
 QY 516 TTGAGATGGCCAGAGTTCTCTGTAAGTCTCTCTTCACTGCTGTCATTAAGCACTAT
 Db 293 CAAGTGGATAGTGAAGCTTTCGTGCAAGAGTTGATTCGAATCCTCAAAATATGTCG 352
 QY 576 AGGCTGCTTAGAGTGGCAAGCTT--TCTCG-GCCA-GAATCTCGAC-TGTC-T--TGA-G 626
 Db 353 CAATATGATATATCAGCGGAATATCGACATTCGACATTTGATAAATTCGGAGCTTTGT 412
 QY 627 CACTGCTGCTCCAGCTCAGCGACAGGCTCTTCTGACTTTGAGCAGCTGGAAGATCCTG 686
 Db 413 GATATGATCCAGATGGAATATATATTTGATGATGCGCAATGGTTATTCGGAGTA 472
 QY 687 GAAGCTATTCCTCCAGGTAAGATATATATGCTGATGCTGCAATGGCTACTGGAACAC 746
 Db 473 TTGCTGACTTTCATACGACGCGTACGAGCAATTTTCCACACATACATCTTTGGCGGG 532
 QY 747 TTGTTGAATTTGTAAGAAAGATGTAAGCAAGCGCTTCCCGACGACATCATGCGAGGG 806
 Db 533 AATGCTGTCAAGGTGAATGTTGAGAGCTCATACTATCCGCGCGGAGCGTGTGAAG 592
 QY 807 AATGTGTATACAGAGAGATGTTAGAAAGCTATCTTCTGGGCGTGAATCATCAAA 866
 Db 593 GTTGGCATCGGTCCGGGGTGGGTGTGACACAGCGGTGAAGAGCTGTGGATACCCA 652
 QY 867 GTGGGAATTTGGCGAGGCTGTGTGTACTACTGGAAGAAACTGGAAGGGGATATCCA 926
 Db 653 CAGCTGATGCGGTGACTGCAATGTGACAGCATCTCATGCTCTCAAGGGGACAGTATG 712
 QY 927 CAGCTGAGCCGATGATGAGTGTGACATGCTCTCATGAGCTCTCAAGGCGACATCATTT 986
 Db 713 AGCAGCGTGGATGTACGAATCTCTGAGAGCGTCTCAAGGCGTGGCGGTGTGAT 772
 QY 987 TCAGATGAGAGTTGCTCAGCTGCTGCGGGAGTGGCCAAAGGCTTTGGGCGAGAGCTGAC 1046
 Db 773 TTGCTTATGATAGAGATTTATTGGCTGGCCATGATAGTGTGGTGAAGTGAAG 832
 QY 1047 TTGCTTATGCTGGGTGCAATGCTGGCTGGGCGACATGATGAGTGTGAGCTCATGAG 1106
 Db 833 AAGATGCGAAGAGTATTAATTTCTATGGATGTCTCTGATAGCGGAATGAAGAA 892
 QY 1107 AGGATGCGCAAGAGTATCAAGCTTTTATGGAATGAGTTCTGAATGGCCATGAAG 1166
 Db 893 TATCAGGATGAGCTTGGCGAATATCTGCTTCTGAGGGCAAAAGATCTACATCGCTTAC 952
 QY 1167 TATGCTGGGGCGTGCTGAGTACAGAGCCTCAGAGGGAAGACAGTGAAGTCTTTT 1226
 Db 953 AGAGTACGATATCGCAAGATACAGATGCTGAGAGGCTTACGCTGGCATGCACC 1012
 QY 1227 AAGAGATGTGGAACATATCCATCCGAGACATCTTAGAGGAGATCCGCTCTACGTAC 1286
 Db 1013 TATATGAGACCACTAGCTCAAAAGACTCTCAAAAAGGCGCAGCTTGTCCGTGACG 1072
 QY 1287 TATGTGGAGACAGCTAAGCTCAAAAGAGTTGAGCGAGGAACCTATCTTCATCCGAGTACC 1346
 Db 1073 CAACAG 1078
 QY 1347 CAGCAG 1352
 RESULT 3
 LOCUS ECAB000119 12434 bp DNA BCT 02-SEP-1997
 DEFINITION Escherichia coli K-12 MG1655 section 9 of 400 of the complete
 genome
 ACCESSION AF000119 U00096
 NID 91786283
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 12434)

AUTHORS
 Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Shao, Y.
 TITLE
 JOURNAL The complete genome sequence of Escherichia coli K-12
 MEDLINE Science 277 (5331), 1453-1474 (1997)
 REFERENCE 97426617
 2 (bases 1 to 12434)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 3 (bases 1 to 12434)
 REFERENCE
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

COMMENT

The E. coli K-12 sequence and its annotations have been updated.
 All of the ambiguous residues in our original submission have been
 resolved, and mis-assemblies in two repetitive regions have been
 realigned. The annotations have been improved and updated as well.
 With this release we begin designating a version number for the
 annotated sequence, to assist in keeping track of corrections,
 updates, and other changes. This is version M52 (SRPT. 02. 1997).
 In addition, a revised notation has been instituted which assigns
 each gene (protein- or RNA-encoding) a unique numeric identifier
 beginning with a lowercase 'b' (in the 'label' field); this will
 remain constant through further updates, gene identifications, etc.
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K-12 strain MG1655.
 Predicted open reading frames were determined using Genemark
 software, kindly supplied by Mark Borodovsky, Georgia Institute of
 Technology, Atlanta, GA, 30332.
 e-mail: mark@ember.gatech.edu
 Open reading frames that have been correlated with genetic loci are
 being annotated with CG Site Nos., unique ID nos. for the genes in
 the E. coli Genetic Stock Center (CGSC) database at Yale
 University, kindly supplied by Mary Berlyn. A public version of the
 database is accessible (<http://cgsc.biology.yale.edu>).
 Annotation of the genome is an ongoing task whose goal is to make
 the genome sequence more useful by correlating it with other data.
 Comments to the authors are appreciated. Updated information will
 be available at the E. coli Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>).

FEATURES

source
 1..12434
 /organism="Escherichia coli"
 /strain="K-12"
 /sub_strain="MG1655"
 /db_xref="taxon:562"
 61..1212
 /gene="ftsZ"
 /gene="ftsZ"
 61..1212
 /note="c383; 99 pct identical to ftsZ_ECOLI SM: P06138"
 /codon_start=1
 /label="b0095"
 /product="cell division protein FtsZ"
 /db_xref="PID:91786284"
 /transl_table=11
 /translation="MEPEMETNDAAVIRKIVGGGAGNAVEHMERIEGVFEAVNT
 DAQALRKTAAGOTIOIGSITKIGAGANPEVRGAADRDRLAALEGAMVRIA
 GNGGCTGTAAPVAFAVAKDGLIVAVYTRFNRGKKRAFAAGITTELSKRVLS
 TITPDKRLKIVGRGISLIDARGANDVKGAVQVIGIAELITRPGMLADFAVNRVMS
 ENGIVAMGSGVAGEDRAEAEMAISSPLEDDIDLSGARGLAVITRGFDLRDEFE
 IVGNITRAFAVDNAATVIVIGTSLDPMDELKRVTVATGIGMDKREPIITLVTKVOQOP

[illegible]

```

/db_xref="PID:g1786288"
/translation="MKRLQIAGVGIIRNENNEIFITRRADAHANAKLEPPGKIEGME
TPEAVYRELQEEVGIPPHQSLFPEKLEYEFPPDRHITLFWLIVEREGWEGQPG
EEMSLVGLNADDEPPANEPVIATKRL"
complement(6330..6454)
/feature=f44; 61 pct identical (1 gap) to 21 residues of an
approx. 216 aa protein KTH_HUMAN SW: P23919"
/codon_start=1
/label-b0100
/db_xref="PID:g1786289"
/translation="MALFPKAMTGAKNQSSDILMPHYGCIIRGQRIRHLVMSDPA
"
complement(6405..6602)
/gene="yacc"
complement(6405..6602)
/gene="yacc"
/feature=f85; 100 pct identical to YACG_ECOLI SW: P36681 (50
aa) but contains 15 additional N-terminal residues"
/codon_start=1
/label-b0101
/db_xref="PID:g1786290"
/translation="MSEITIVNCPPTGKTVWGEISPFPCSKRCQLIDLGMAEE
KRIPLSGDLSDDMSEPRQ"
complement(6612..7355)
/gene="yacc"
complement(6612..7355)
/gene="yacc"
/feature=f247; 98 pct identical to YACF_ECOLI SW: P36680 (78
aa) but has 169 additional N-term residues"
/codon_start=1
/label-b0102
/product="hypothetical protein in mutr-guac intergenic
region"
/db_xref="PID:g1786291"
/translation="MQRQVLFEPHLEPKMRTWLRIEFLIQQLNVNLPYDHAALHFF
RNVSELDVFERGEVRETLKELDRQRKIQTMVGVGVDSQREALIQKRAAGSVL
ISARIGQFLREDEPLILVRORLSIRPGCCSPFLPHITLHLRQMRDSQVETWIAS
LNPTQALTNVLDIROSAPFRKQTSLNQFYQDNQGDADLLRLNLSLDOLYPOISGH
KSRAIRFMPDTENGVPERLDPELACC"
Note: remainder of annotations omitted.

Query Match 11.0%; Score 203; DB 13; Length 12434;
Best Local Similarity 60.6%; Pred. NO. 1.02e-153;
Matches 623; Conservativeness 0; Mismatches 402; Indels 3; Gaps 3;

Db 8205 TATTGAGAGATCTGAGATTAGGTTTAAAGAGCTTCTATCCGCCCTAAAGCTCCAC 8264
QY 344 TATTACACACGATGTGAACCTGACTTAAAGAGATGCTCTTTGAGGCCCAACGACAGTAC 403
Db 8265 TCTTAAAAACCGCTTCGATGTGAACCTGGAACGTCATTTCACTTCAACATTTGAGGTCA 8324
QY 404 CCTTAAGCTCTGGAAGTAGAGTGAGATCTCAACAAGATCCCTTTCAATTGCGAATCAAGCA 463
Db 8325 GAGCTGCTCCGGCGCTGCGGATTAATGCGCGCAAAATATGACACCGTGAACATTTTCTAT 8384
QY 464 GACATACCTCTGGGGTTCCTCATCATGTGCTGCCAAATATGATGATCTGCGGACACTTGGAGAT 523
Db 8385 GGCCCTCGCGGCTGTTTGTGATTTTGAATTTTACAGTGCATGATTAACACATTTCTGTCGA 8444
QY 524 GGCCAGGTTCTCTGATTTCTCTCTCTTTACAGTGCATGATTAACACATTTTCTGTCGT 583
Db 8445 AGAGTGGCAAGCGTTTATCAACAATTTCTTCGCTGATGTGCTGAACAATGATGATGTTTC 8504
QY 584 TCAGTGGCAAGGTTTGTGCGCAAGATTCGACGTGTTGAGATC-TG-GCTGCGACG 641
Db 8505 TACCGGTACGTCTGATGCGGATTTGAAAAAATCAACAGATTTCCGACCTGAACCCGGC 8564
QY 642 T-CAGGCAACAGGCTTCTTGTGACTTTGACACACTGGAAACAGATCCCTGGAAAGCTATTATCCCA 700

```

Db	8565	ATTAACTTGTTGTTGATTTGACTGGCGAATGGTTATTTCGCCAACCCTTGTGCGCACTTGT	8624
QY	701	GGTGAATATATATATGCGTGGATGTGGCAATATGGCTATCTGTGAACACTTGTGTAATTTGT	760
Db	8625	TGCGAAGACCCGCGTGAAGCGCGGCGCGACCAAAACCATTTGTGCTGGTAACGTAGTACTGG	8684
QY	761	AAAAGATGTACGGAAGCGCTTCCCCAGCAGCACACATCATGCGAAGGAAATGTGTAAACAG	820
Db	8685	TGAATTTGTGAGAGACTTATCTCTCAGGTGCCATATCGTTAAAGTTGGCATTGGCCC	8744
QY	821	AGAAATGTTAGAAAGAGCTCATCTTCTTGCGGCTGCATCATCAAAAGTGGAAATTTGGCC	880
Db	8745	AGGTTCTGTTTATCAACTGCGCTCAAAACAGCGCGGTTATTCGGCAACTTCTGCGGT	8804
QY	881	AGGCTCTGTGTGTACTACTGGAAAGAAATGGAGTGGGGTATCCACAGCTCAGCGCACT	940
Db	8805	AATGGAATGTGCGCATGCTCGCACGGTCTGGGCGGAATGATCGTACAGATGTGGCGTG	8864
QY	941	GATGAGATGTGCAAGATCTCTCATGGCTCAAAAGGCCATCATTTTCAGATGAGGTTG	1000
Db	8865	CACCACCGCGGCGCATGTGCGAAAGCCTTTGGCGGCGGTGCCATTTTCGTACGTTGG	8924
QY	1001	CAGGTCTCTGGGGATGTGGCCAAAGCCTTTTGGGCGAGAGCTGACTTCGTATCTGGG	1060
Db	8925	CGGATCTGTGGCGGCCACGAAAGAGAGCGGCGTGCATCTGTTGAGAGAAACGCGGCAAA	8984
QY	1061	TGCGATCTGTGCTGGCGACAGTGAAGTGGTGTGACTCATCGAGAGAGATGGCAAGAA	1120
Db	8985	ATTATCTGTTCTACGCGCATGTAGCTCCGAGTCTGCGATGAAGAAACGTCAAGTTCGGGCGT	9044
QY	1121	GTACAGACTCTTCTATGATGATGAGTTCTGAATATGGCCATGAAAGATATGCTGGGGGGT	1180
Db	9045	TGCGGAATATTCGCGCAGCAGAAAGTTAAACCGTTAAGTCTGCCGTGCGAGGCCGTTGA	9104
QY	1181	GGCTGAGTACAGAGACCCACAGAGGAAAGACAGTGAAGTTCCTTTAAAGGATGTGA	1240
Db	9105	AAATACCGCGGAGATTTTTTGGCGGCGCTGCGTTCAGTGTACATACGTTGGGGCTTC	9164
QY	1241	ACATACCATCCGAGACATCTCTAGAGAGGATCGCTTACGTGTACCTATGTGGGAGCG	1300
Db	9165	ACGCGTGAAGAGCTGACCAAGCGCACCAAGTTATTCTGTGCGAGAACAGAAAACCG	9224
QY	1301	TAACTCAAGAGTTGAGCAGAGAACTACCTCATCCGATCCAGATCACCCAGAGTGTAATCC	1360
Db	9225	CATCTCA 9232	
QY	1361	AATCTTCA 1368	
RESULT	4		
LOCUS	ECGUACGSM	1991 bp	DNA
DEFINITION	Escherichia coli	guac gene for guanosine 5'-monophosphate	03-DEC-1992
ACCESSION	NID	X07917 M33020	
KEYWORDS	ORGNISM	gmp reductase; guac gene; guanosine 5'-monophosphate	
SOURCE	ORGNISM	Escherichia coli.	
REFERENCE	AUTHORS	1 (bases 1 to 1991)	
TITLE	JOURNAL	Submitted (15-JUN-1988) Andrews S.C., The Dept of Microbiology,	
		Sheffield University, Western Bank, Sheffield, S10 2TN, United	
		Kingdom	
		2 (bases 1 to 1991)	
		Andrews, S.C. and Guest, J.R.	
		Nucleotide sequence of the gene encoding the gmp reductase of	
		Escherichia coli K12	
		Biochem. J. 235 (1), 35-43 (1988)	

MEDLINE 89061679		Location/Qualifiers
FEATURES		Source
-35_signal	repeat_unit	1..1991 /organism="Escherichia coli" /strain="K12" /db_xref="taxon:562" /clone="pGS89" /map="2.6 min" 25..30 /note="Promoter 2" 44..49 /note="inverted repeat A" 50..55 /note="Promoter 2" 61..>1250 /note="transcript from promoter 2" 62..67 /note="inverted repeat A'"
-35_signal	repeat_unit	98..103 /note="Promoter 1" 119..130 /note="inverted repeat B" 122..127 /note="Promoter 1" 126..131 /note="Promoter 3" 128..134 /note="discriminator sequence promoter 1" 132..147 /note="inverted repeat B'"
-10_signal	repeat_unit	135..>1250 /note="transcript from promoter 1" 148..153 /note="Promoter 3" 155..161 /note="discriminator sequence promoter 3" 160..>1250 /note="transcript from promoter 3" 198..201 /note="ribosome binding site" 210..1250 /codon_start=1 /product="GMP reductase (AA 1-346)" /db_xref="PID:g42616" /db_xref="SWISS-PROT:P15344" /transl_table=1 /translation="MRTEEDLKGFKPYLIIRPKRSTLKSRSVDLEKQTFPHKSGQSM SGVITIANMDVTGTFESMASLAFDILTAVHKHYVEEWOAFINNSADVLKRVMS TGTSDADFEKTKQILDLNPALNFCIDVANGYSEHFVOFAKAREAMPRTKICGNVVS TGECEBELILSGADIVKVGIGPGSVCTRYATGVTGVPOLSAVICADAAHGLIAMS DGGCTTPGDVAKAFARADFYVLGGLAGHESGGRIYEENGEKPLFYGSSSEAMKR HVGVAIVRAAEKGTVALPLRGVPENTARDILGSLRSACTYVGASRLKELIKRTTIFIR VOEDENRIFNNL"
-35_signal	repeat_unit	285..291 /note="inverted repeat C" 301..307 /note="inverted repeat C'"
-10_signal	repeat_unit	620..627 /note="inverted repeat D" 644..651 /note="inverted repeat D'"
repeat_unit	repeat_unit	908..915 /note="inverted repeat E" 923..928 /note="inverted repeat F" 932..939 /note="inverted repeat E'"
repeat_unit	repeat_unit	933..938 /note="inverted repeat F'"
repeat_unit	repeat_unit	979..987 /note="inverted repeat G" 997..1005 /note="inverted repeat G'"

```

repeat_unit      1050..1055
                  /note="inverted repeat H"
repeat_unit      1064..1069
                  /note="inverted repeat H"
terminator       1257..1286
                  /note="pot. transcriptional terminator"
repeat_unit      1324..1329
                  /note="inverted repeat J"
repeat_unit      1335..1340
                  /note="inverted repeat J"
repeat_unit      1569..1577
                  /note="inverted repeat J"
repeat_unit      1598..1606
                  /note="inverted repeat K"
repeat_unit      1710..1715
                  /note="inverted repeat K"
repeat_unit      1733..1738
                  /note="inverted repeat L"
repeat_unit      1794..1802
                  /note="inverted repeat L"
repeat_unit      1810..1818
                  /note="inverted repeat M"
repeat_unit      1810..1818
                  /note="inverted repeat M"
BASE COUNT      493 a 494 c 508 g 496 t
ORIGIN

```

```

Query Match      10.1%; Score 186; DB 13; Length 1991;
Best Local Similarity 60.5%; Pred. No. 7,10e-138;
Matches 622; Conservative 0; Mismatches 400; Indels 6; Gaps 5;

```

```

Db 215 TATGAGACAGATCTGAAGTATGATTTTAAAGACGTTCTCATCCGCTAAACGCTCCAC 274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 344 TATGAGACAGATCTGAAGTATGATTTTAAAGACGTTCTCATCCGCTAAACGCTCCAC 403
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275 TCTTAAACCCGCTTCGATGTTGAACGTGAACATTCACCTTCAACATTCAGGTCA 334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 404 CCTTAAAGCTCGAAGTGTGAGTGCATCAAAAGATCTTTTCATTCGGAACCTCAAGCA 463
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 335 GAGCTGCTCCGCGCGGATTTATCGCCGCAAAATATGACACCCGACGACATTTTCAT 394
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 464 GACATACCTCTGGGTTCCCATCATTTGCGCAATATGATGCTGGGACCTTTGAGAT 523
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 395 GGCCTCTCGCTGGCTTTTATGATTTTGAATTTGCTGCTGATTAACACTATTCGTGCA 454
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 524 GGCACAGTTCTCTGTAGTCTCTCTCTCACTCTGTCCATTAAGCACTAATAGCTCTGT 583
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 AGAGTGCAGACGTTTATCAACAATTCCTCGCTGATGCTGAAACATGTGATGTTTC 514
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 584 TCAGTGGCAAGATTTGCTGGCCAAATCTGATGCTGTGAGCATC-TG-GCTCCAGC 641
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 515 TACGGGTACGCTGATGCGGATTTGCAAAAAAACAATTCGACCTGACACCCGCG 574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 642 T-CAGGCGACAGGCTCTTCTGACTTTGACAGCTGGAAAGATCCTGGAAGCTATTCGCCA 700
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 575 ATTAAACTCTGTTGTATTTGACGTGGCGATGTTATTCGGAACCTTCGTGACATTCGT 634
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 701 GGTGAAGATATATGCCCTGATGTGGCAATGTGCTACTCTGAACCTTTGTTGAATTTGT 760
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 635 TGCAGAAACCGGTGAAGCGGTGGCCGACCAAAACATTTTGCTGTGTAACGTAAGTACTGG 694
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 761 AAAAGATGTAGGAAGCGCTTCCGCCACACACCATTCATGGAGGGAATGTGTAAACAG 820
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 695 TGAATATGTATGAGAGCTTATCTCTCAGGTGCCGATATCGTTAAAGTTGGATGGCC 754
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 821 AGAGATGTGTAAAGAGCTATCTTTCTGGGGCTGACATCATCAAAAGTGGGAATGGGCC 880
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 755 AGGTCTCTGTTTACACATCGCGGTCAAAACAGCGCTGATTATCCGCACTTTCTGGCGT 814
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 881 AGCCTCTGTGTACTACTCGGAAGAAAACGTGAAGTGGGATTCACAGCTCAGCGCACT 940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 815 AATGGAATGTGCCGATGCTGGCGACAGCTGTGGCGGAATGATCGTACAGCATGTGGCTG 874
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 941 GATGAGATGTGCAGATGCTGCTCATGGCCTCAAAAGGCCACATCATTCAGATGAGAGTTG 1000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 875 CACCACGGCGGCGATGAGCGGAAGCGCTTGGC-C--GTGCCGATTTTCGTCATCTGG 931
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1001 CAGCTGTCTTGGGATGTGGCCAAAGCGCTTTGGGCGAGAGCTACTCTTGATGCTGGC 1060
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 932 CGGATGCTGGCGGCGCAAGAGAGCGCGGCTCGCATCGTTGAGGAGAACGCGGAGAA 991
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1061 TGGCATGCTGGCTGGGACAGTGAAGTGAAGTGTGAGTGTATCGAAGAGGATGGCAAGAA 1120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 992 ATTATGCTGTCTACAGGCACTGACCTCGAGTCTGCGATGAACGTCACGTTGGCGGCT 1051
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1121 GTACAGCTCTTATGGAAGAGTTCGAAATGCCATGAAGAGTATGCTGGGGCGT 1180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1052 TCGGGAATATGCGCAGAGAGAGTTAAACCGTTAACTGCGCGCTGGAGCCCGGTTGA 1111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1181 GCGTGAATACAGACCTTCAGAGGAAGACAGTGAAGTCTTTTAAAGAGATGTGA 1240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1112 AAATACCGCGCGAGATATTTTGGCGCGCTGCTTACGTTGATCATATACGTTGGGCTTC 1171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1241 ACATACCATCCGAGACATCCATAGAGAGGATCCGCTACAGCTACATATGTGGAGCAGC 1300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1172 ACGCCTGAAGAGCTGACCAAGCCGACGCTTATTCGTGTGACGAGAACAGAAACCG 1231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1301 TAACTCAAAAGATTTGAGCAGAGAGACTACCTCATCGAGTCAACCCAGAGGTGAATCC 1360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1232 CATCTCA 1239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1361 AATCTCA 1368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT          5
LOCUS           EC082K      82727 bp      DNA      BCT      23-JAN-1997
DEFINITION      Escherichia coli genome, 2.4-4.1 mln region (110,917-193,643 bp
                from 0 min).
ACCESSION      D26562
NID            9473770
KEYWORDS        7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase; GMP reductase;
                pili uridylyl-transferase; adenosylmethionine decarboxylase
                proenzymes; ampd signalling protein; ampe signalling protein;
                aromatic amino acid transport protein; deoxyguanosine triphosphate
                triphosphohydrolase; dihydroliipoamide acetyltransferase;
                dihydroliipoamide dehydrogenase; dosage-dependent dnaK suppressor
                protein; elongation factor EF-Ts; ferrichrome-iron receptor;
                ferrichrome-iron transport protein fhbB; ferrichrome-iron transport
                protein fhuc; ferrichrome-iron transport protein fhnd; glucose
                dehydrogenase; glutamate-1-semialdehyde aminotransferase; heat
                shock protein htra; methionine aminopeptidase; mukB suppressor
                protein smba; mutator mutT; nicotinate-nucleotide pyrophosphorylase
                (carboxylating); penicillin-binding protein 1B; plasmid copy number
                control protein pcnB; pyruvate dehydrogenase; regulatory protein
                sfsl; ribosomal protein S2; ribosome releasing factor; spermidine
                synthase; tetrahydrodipicolinate N-succinyltransferase.
                Escherichia coli (sub-strain W3110, strain K-12) (library: Kohara's
                mini set library) DNA, clone #111, 112, 113, 114, 115, 116 and 119.
                Escherichia coli
ORGANISM        Escherichia; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
SOURCE          An.G., Bendlak,D.S., Mamelak,L.A. and Friesen,J.D.
                Organization and nucleotide sequence of a new ribosomal operon in
                Escherichia coli containing the genes for ribosomal protein S2 and
                elongation factor Ts
REFERENCE       Nucleic Acids Res. 9 (16), 4163-4172 (1981)
JOURNAL        MEDLINE
AUTHORS        Stephens,P.E., Darlison,M.G., Lewis,H.M. and Guest,J.R.
TITLE          The pyruvate dehydrogenase complex of Escherichia coli K12.
                Nucleotide sequence encoding the pyruvate dehydrogenase component
                Eur. J. Biochem. 133 (1), 155-162 (1983)
JOURNAL        MEDLINE
AUTHORS        Stephens,P.E., Darlison,M.G., Lewis,H.M. and Guest,J.R.
TITLE          The pyruvate dehydrogenase complex of Escherichia coli K12.
                Nucleotide sequence encoding the dihydroliipoamide acetyltransferase

```

- JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
component
Eur. J. Biochem. 133 (3), 481-489 (1983)
83234434
4 (sites)
Stephens, P.E., Lewis, H.M., Darlison, M.G. and Guest, J.R.
Nucleotide sequence of the lipamide dehydrogenase gene of
Escherichia coli K12
Eur. J. Biochem. 135 (3), 519-527 (1983)
84004369
5 (sites)
Richaud, C., Richaud, F., Martin, C., Haziza, C. and Patte, J.C.
Regulation of expression and nucleotide sequence of the Escherichia
coli dcpd gene
J. Biol. Chem. 259 (23), 14824-14828 (1984)
85054973
6 (sites)
Broome-Smith, J.K., Edelman, A., Yousif, S. and Spratt, B.G.
The nucleotide sequences of the ponA and ponB genes encoding
penicillin-binding protein 1A and 1B of Escherichia coli K12
Eur. J. Biochem. 147 (2), 437-446 (1985)
85127060
7 (sites)
Coulton, J.W., Mason, P., Cameron, D.R., Carmel, G., Jean, R. and
Rode, H.N.
Protein fusions of beta-galactosidase to the ferrichrome-iron
receptor of Escherichia coli K-12
J. Bacteriol. 165 (1), 181-192 (1986)
86085668
8 (sites)
Breton, R., Sanfacon, H., Papayannopoulos, I., Biemann, K. and
Lapointe, J.
GlutamyI-tRNA synthetase of Escherichia coli. Isolation and primary
structure of the gltX gene and homology with other aminoacyl-tRNA
synthetases
J. Biol. Chem. 261 (23), 10610-10617 (1986)
86278132
9 (sites)
Koster, W. and Braun, V.
Iron hydroxamate transport of Escherichia coli: nucleotide sequence
of the hnf gene and identification of the protein
Mol. Gen. Genet. 204 (3), 435-442 (1986)
87014116
10 (sites)
Chye, M.L. and Pittard, J.
Transcription control of the aroP gene in Escherichia coli K-12:
analysis of operator mutants
J. Bacteriol. 169 (1), 386-393 (1987)
87083395
11 (sites)
Ben-Bassat, A., Bauer, K., Chang, S.Y., Myambo, K., Boosman, A. and
Chang, S.
Processing of the initiation methionine from proteins: properties
of the Escherichia coli methionine aminopeptidase and its gene
structure
J. Bacteriol. 169 (2), 751-757 (1987)
87109068
12 (sites)
Coulton, J.W., Mason, P. and Allatt, D.D.
hnf and hnfD genes for iron (III)-ferrichrome transport into
Escherichia coli K-12
J. Bacteriol. 169 (8), 3844-3849 (1987)
87279948
13 (sites)
Tabor, C.W. and Tabor, H.
The speB gene of Escherichia coli. Formation and processing
of a proenzyme form of S-adenosylmethionine decarboxylase
J. Biol. Chem. 262 (33), 16037-16040 (1987)
88058963
14 (sites)
Gebhard, W., Schreitmüller, T., Hochstrasser, K. and Wachter, E.
Complementary DNA and derived amino acid sequence of the precursor
of one of the three protein components of the inter-alpha-trypsin
inhibitor complex
- JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
FEBS Lett. 229 (1), 63-67 (1988)
88152237
15 (sites)
Mellano, M.A. and Cooksey, D.A.
Nucleotide sequence and organization of copper resistance genes
from Pseudomonas syringae pv. tomato
J. Bacteriol. 170 (6), 2879-2883 (1988)
88227880
16 (sites)
Andrews, S.C. and Guest, J.R.
Nucleotide sequence of the gene encoding the GMP reductase of
Escherichia coli K12
Biochem. J. 255 (1), 35-43 (1988)
89061679
17 (sites)
Sung, Y.C. and Fuchs, J.A.
Characterization of the cyn operon in Escherichia coli K12
J. Biol. Chem. 263 (29), 14769-14775 (1988)
8908347
18 (sites)
Lipinska, B., Sharma, S. and Georgopoulos, C.
Sequence analysis and regulation of the htrA gene of Escherichia
coli: a sigma 32-independent mechanism of heat-inducible
transcription
Nucleic Acids Res. 16 (21), 10053-10067 (1988)
89057448
19 (sites)
Liu, J.D. and Parkinson, J.S.
Genetics and sequence analysis of the pcnB locus, an Escherichia
coli gene involved in plasmid copy number control
J. Bacteriol. 171 (3), 1254-1261 (1989)
89155419
20 (sites)
Xie, Q.W., Tabor, C.W. and Tabor, H.
Spermidine biosynthesis in Escherichia coli: promoter and
termination regions of the speD operon
J. Bacteriol. 171 (8), 4457-4465 (1989)
89327165
21 (sites)
Lindquist, S., Galleni, M., Lindberg, F. and Normark, S.
Signalling proteins in enterobacterial Ampc beta-lactamase
regulation
Mol. Microbiol. 3 (8), 1091-1102 (1989)
90113890
22 (sites)
Bonero, M.I., Jepsen, L.P., Stroman, P. and van Heeswijk, R.
Characterization of a leuA gene and an ARS element from Mucor
circinelloides
Gene 84 (2), 335-343 (1989)
90128278
23 (sites)
Honore, N. and Cole, S.T.
Nucleotide sequence of the aroP gene encoding the general aromatic
amino acid transport protein of Escherichia coli K-12: homology
with yeast transport proteins
Nucleic Acids Res. 18 (3), 653 (1990)
90174991
Erratum: [Nucleic Acids Res 1990 Mar 11;18(5):1332]
24 (sites)
Surin, B.P., Watson, J.M., Hamilton, W.D., Economou, A. and Downie, J.A.
Molecular characterization of the nodulation gene, nodT, from two
biovars of Rhizobium leguminosarum
Mol. Microbiol. 4 (2), 245-252 (1990)
90251164
25 (sites)
Kang, P.J. and Craigh, E.A.
Identification and characterization of a new Escherichia coli gene
that is a dosage-dependent suppressor of a dnaK deletion mutation
J. Bacteriol. 172 (4), 2055-2064 (1990)
90202727
26 (sites)
Wurgler, S.M. and Richardson, C.C.
Structure and regulation of the gene for dGTP triphosphohydrolase

Note: remainder of annotations omitted.

```

Query Match      10.0%; Score 185; DB 15; Length 39235;
Best Local Similarity 71.8%; Pred. No. 6,03e-137;
Matches 305; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Db 9811 TTCGCACTCAAGACATCTCTTGAAGTGGCCAGGAGATCTCAATCTTCGTGG 9870
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 694 TTCCCAAGTGAATATATATGCTGATGTCGCAATGGCTACTTAACACTTTGTG 753
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9871 AGTTATTCGCGCTGTTCGAGAACCTACCCGAGACACACATCATGGCCGGAACGTG 9930
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 754 AATTGTAAAGATGTAGGGAAGCGCTTCCCCAGACACACATCATGGCAGGAGATG 813
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9931 TCACCGAGAGATGTTGAGAGCTGATTTCTGTGCGCGCACATTTGCAAGTGGAA 9990
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 814 TAACAGGAGAGATGTTAGAGAGCTCATCTTTCTGGGCTCATCATCAAGAGTGGAA 873
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9991 TCGGACCGGANTCGGTTTGACACCACTCCGCAAGAGCCGAGATTGGATACCAACGCTTA 10050
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 874 TTGGCCAGGCTCTGTGTGTACTACTCGGAAGAAAAGTGGATGCCATCCACAGCTCA 933
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10051 GTCCGCTCTGAGAGTCCGCTGATGCCGCTCATGTCTCATGAGATGATGATG 10110
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 934 GCGCAGTATGAGATGTGTGACAGATGCTGCTCATGGCTCAAGGCCACATCATTTAGATG 993
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10111 GTGATGACAGCAATCTTGAGATGTTGCCAAGGCTTTTGAGAGCTGAGCAGATTTGCTCA 10170
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 994 GAGGTTCGAGCTGCTCTGGGATGTGCCAAGGCTTTTGAGGAGAGCTGACTCTGCTGA 1053
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10171 TGATTCGAGAGACTCTTCCTGACACGATCAGATGAGGAGAGATCTCATGAGCATATG 10230
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1054 TGTGTGGTGGCTGCTGCTGGGACAGTGAAGTGTGAGTCAATCGAGAGGATG 1113
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10231 GAAG 10235
      |||||
Qy 1114 GCAAG 1118
      |||||

RESULT 7
LOCUS S73035S9 601 bp DNA PRI 10-JUL-1992
DEFINITION guanosine monophosphate reductase [human, Genomic, 601 nt, segment 9 of 9].
ACCESSION S73075
NID 9241139
KEYWORDS 9 of 9
SEGMENT human.
SOURCE Homo sapiens
ORGANISM Eukaryotes; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 601)
AUTHORS Kondoh,T., Kanno,H., Chang,L. and Yoshida,A.
TITLE Genomic structure and expression of human guanosine monophosphate
reductase
JOURNAL Hum. Genet. 88 (2), 219-224 (1991)
MEDLINE 92098099
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI g1bseq 73075] from the original journal article.
This sequence comes from figure 5.

FEATURES
SOURCE
1. .601
Location/Qualifiers
/db_xref="taxon:9606"
/organism="Homo sapiens"
order(S73035:1..279,S73043:1..159,S73045:1..130,
S73047:1..232,S73049:1..158,S73053:1..170,S73060:1..96,
S73066:1..207,1..601)
/gene="guanosine monophosphate reductase"

BASE COUNT 131 a 167 c 140 g 163 t

Query Match 5.1%; Score 94; DB 22; Length 601;

```

```

Best Local Similarity 76.1%; Pred. No. 2.13e-54;
Matches 137; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Db 36 CAGAGCCTCTGAGGGAAGACTGTGGAAGTTCCTTACAAAGAGATGTGAAACACTAT 95
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1190 CAGAGCCTCAGAGGAAGAACAGTGAAGTTCCTTTAAAGAGATGTGAAATACAT 1249
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96 CCTGATATTTCTGGGGACAGTCCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 155
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1250 CCGAGACATCTTACAGAGGATCCCTCTGATATGTTGGACACTTAAGCTCAA 1309
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156 GGACCTCAGAGAGGCAACATTCATCCGGTGCAGCAGCAGCAGCAACACCGTTCAG 215
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1310 AGATTGAGCAGGAGACTTCTTATCCAGTACCCAGCAGGTGAATTCATCTTCAG 1369
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
LOCUS HS467D16 143583 bp DNA PRI 18-JAN-1998
DEFINITION Human DNA sequence from PAC 467D16 on chromosome 6p22.3-24.1.
Contains the 3' part of the SCAL (ataxin-1) gene with a
poly-glutamine (CAG repeat) polymorphism, the 3' part of the GMPR
(GMP reductase, Guanosine 5'-monophosphate oxidoreductase) gene,
ESTs and an STS with a polymorphic CA repeat.
AL009031
6p22.3-24.1; ataxin-1; CA repeat polymorphism; CAG repeat
polymorphism; GMPR; Guanosine 5'-monophosphate oxidoreductase;
SCAL.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 143583)
Tubby,B.
REFERENCE Submitted (19-JAN-1998) sanger.ac.uk/HGP/Chrf6/ Sanger Centre,
Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
IMPORTANT: This sequence is the entire insert of clone 467D16.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre chromosome 6
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrf6/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key
The true left end of clone 467D16 is at 1 in this sequence. The
true right end of clone 467D16 is at 143583.
467D16 is from the library RBC13 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.

FEATURES
SOURCE
1. .143583
Location/Qualifiers
/db_xref="taxon:9606"
/organism="Homo sapiens"
/chromosome="6"
/map="6p22.3-24.1"
/clone="467D16"
/clone_id="RBC13"
99. .252
repeat_region
/note="MER20 repeat: matches 164. .4 of consensus"
prim_transcript complement(507..931)
/note="match: ESTs T05971"
repeat_region 1860..1929

```

```
repeat_region /note="MIR repeat: matches 48. .116 of consensus"
2345. .2606 /note="AluSx repeat: matches 34. .302 of consensus;
incomplete repeat"
repeat_region /note="MER7B repeat: matches 1. .618 of consensus"
3454. .3575 /note="AluJo repeat: matches 124. .3 of consensus;
incomplete repeat"
repeat_region /note="MER7B repeat: matches 645. .885 of consensus"
3604. .3841 /note="MER7B repeat: matches 299. .1 of consensus"
3842. .4139 /note="AluSc repeat: matches 866. .1204 of consensus"
4140. .4481 /note="MER7B repeat: matches 303. .47 of consensus;
incomplete repeat"
repeat_region /note="AluSg repeat: matches 867. .1042 of consensus"
5280. .5454 /note="match: ESTs AA432016 AA431766"
5413. .5956 /note="AluSg repeat: matches 122. .1 of consensus;
incomplete repeat"
repeat_region /note="AluJo repeat: matches 11. .295 of consensus"
6709. .7049 /note="match: EST AA320301"
6908. .6964 /note="MER5A repeat: matches 180. .123 of consensus"
6968. .7048 /note="MER5A repeat: matches 80. .1 of consensus"
7080. .7376 /note="AluSg repeat: matches 1. .298 of consensus"
8333. .8541 /note="MER5B repeat: matches 2. .178 of consensus"
8694. .8996 /note="AluSx repeat: matches 302. .1 of consensus"
9157. .9194 /note="19 copies of 2 mer 92 & conserved"
complement(9168. .9193)
/note="13 copies of CA 100 & conserved; differs from
216874"
repeat_region /note="AluSg repeat: matches 300. .1 of consensus"
10519. .10820 /note="MIR2 repeat: matches 146. .41 of consensus"
12763. .12864 /note="AluJo repeat: matches 1. .132 of consensus;
incomplete repeat"
repeat_region /note="MSRA repeat: matches 426. .1 of consensus"
14964. .15339 /note="MIR30 repeat: matches 116. .1 of consensus"
17034. .17149 /note="AluJo repeat: matches 300. .85 of consensus;
incomplete repeat"
repeat_region /note="AluSg repeat: matches 298. .1 of consensus"
18572. .18865 /note="MIR1C repeat: matches 186. .1 of consensus"
18866. .19049 /note="MIR2 repeat: matches 98. .144 of consensus"
19568. .19613 /note="MIR2 repeat: matches 62. .146 of consensus"
20206. .20294 /note="MIR2 repeat: matches 898. .367 of consensus"
20468. .20984 /note="L1PB1 repeat: matches 3. .302 of consensus"
22600. .22841 /note="MIR1A2 repeat: matches 374. .128 of consensus"
26002. .26363 /note="THERC repeat: matches 371. .1 of consensus"
26942. .27243 /note="AluJo repeat: matches 302. .1 of consensus"
27997. .28318 /note="AluJo repeat: matches 1. .300 of consensus"
29651. .29808 /note="MIR repeat: matches 62. .217 of consensus"
30254. .30403 /note="match: ESTs T19981 T19980"
30465. .30582 /note="2 copies of 59 mer 86 & conserved"
31248. .31452 /note="match: EST AA364443"
31686. .31946 /note="AluSx repeat: matches 302. .44 of consensus;
incomplete repeat"
repeat_region /note="3 copies of 22 mer 94 & conserved"
32473. .32512 /note="20 copies of 2 mer 95 & conserved"
33075. .33376 /note="AluSx repeat: matches 302. .1 of consensus"
34397. .34699 /note="AluSx repeat: matches 1. .301 of consensus"
34734. .34804 /note="L1PA5 repeat: matches 820. .890 of consensus"
35273. .35564 /note="AluY repeat: matches 1. .294 of consensus"
35546. .36664 /note="MIR2 repeat: matches 146. .24 of consensus"
39656. .39703 /note="MER5B repeat: matches 53. .5 of consensus"
39803. .40042 /note="L1PB1 repeat: matches 902. .656 of consensus"
40000. .40058 /note="L1PA7 repeat: matches 698. .640 of consensus"
41306. .41743 /note="L1PB2 repeat: matches 658. .204 of consensus"
42306. .42607 /note="AluSx repeat: matches 1. .302 of consensus"
42615. .42664 /note="5 copies of 10 mer 90 & conserved"
44012. .44314 /note="AluSx repeat: matches 1. .301 of consensus"
45328. .45619 /note="AluY repeat: matches 3. .301 of consensus"
45721. .45874 /note="MER5B repeat: matches 178. .3 of consensus"
45886. .45964 /note="MER5A repeat: matches 188. .109 of consensus"
45887. .45985 /note="MER5A repeat: matches 17. .109 of consensus"
46182. .46482 /note="AluSg repeat: matches 300. .1 of consensus"
46671. .47200 /note="MER44B repeat: matches 536. .1 of consensus"
48173. .48469 /note="AluJo repeat: matches 1. .297 of consensus"
48772. .49074 /note="AluJo repeat: matches 297. .1 of consensus"
49573. .49680 /note="MIR repeat: matches 183. .75 of consensus"
50354. .50652 /note="AluSx repeat: matches 299. .1 of consensus"
50772. .50887 /note="MIR repeat: matches 227. .101 of consensus"
```

'''
Note: remainder of annotations omitted.


```

Note: remainder of annotations omitted.

Query Match          5.1%; Score 94; DB 22; Length 148750;
Best Local Similarity 76.1%; Pred. No. 2,13e-54;
Matches 137; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Db 114946 CTGACAGCGGTGTGTGCCTGGTCCACC GGATGAATTGTTCCCTCCTGAGCTCC 115005
||||| | | | ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db Cp 1369 CTAAGAATTGAAATCACTGCTGGTGACTCGGATGAGGATGTTCTTCCTCTCAACTC 1310
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 115006 TTGAGTTGGCGGCCCCCACAGTAGTGKACACGTGAGCATCACGCCCCCAGAAATATCAGG 115065
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db Cp 1309 TTTGAGCTTACTCTCTCCACATGATGATGCACGTAGACGGATCCCTCTAAGATGTCTCG 1250
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 115066 ATATGTTTTCCACATCTCCTTTGTAAGAACATTCACAGTCTTACCCCTGAGAGCTCTG 115125
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db Cp 1249 ATGATATTCTCCACAATCTCCTTTAAAAAGAACTTCACTGTCTTCCTCTGAGGCTCTG 1190
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10 HSCU177E8 40822 bp DNA PRI 27-JUL-1996
LOCUS Human DNA sequence from cosmid CU177E8, between markers DXS366 and
DEFINITION DXS87 on chromosome X.
ACCESSION Z68694 g1160468
NID x.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 40822)
AUTHORS Williamson,H. and Wilkinson,P.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1995) Sanger Centre, Hinton, Cambridgeshire,
CB10 1RQ, England. E-mail enquiries: humpdes@sanger.ac.uk
COMMENT IMPORTANT: This sequence is not the entire insert of clone
CU177E8. It may be shorter because we only sequence overlapping
sections once, or longer because we arrange for a small overlap
between neighbouring submissions.
The true left end of clone CU177E8 is at 1 in this sequence. The
true right end of clone CU177E8 is at 40822.
The true right end of clone V693A8 is at 34707.
CU177E8 is from the human chromosome X-specific cosmid library.
location/Qualifiers
FEATURES
SOURCE
1..40822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="CU177E8"
2..82
/partial
/note="Alu repeat: matches 210..291 of consensus"
88..119
/note="8 copies of 4 mer 100 % conserved"
88..123
/note="6 copies of 6 mer 94 % conserved"
88..121
/note="17 copies of 2 mer 100 % conserved"
304..383
/note="1L element fragment"
1367..1420
/note="3 copies of 18 mer 91 % conserved"
1367..1414
/note="8 copies of 6 mer 100 % conserved"
1367..1414
/note="12 copies of 4 mer 100 % conserved"
1367..1414
/note="24 copies of 2 mer 100 % conserved"
4399..5185
/note="match: multiple ESTs"
6535..6564
repeat_region
```

```

Note: remainder of annotations omitted.

Query Match          5.1%; Score 94; DB 22; Length 148750;
Best Local Similarity 76.1%; Pred. No. 2,13e-54;
Matches 137; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Db 114946 CTGACAGCGGTGTGTGCCTGGTCCACCAGCATGAATGTTCCCTCCTGAGCTTC 115005
||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | |
Db 115006 TTGAGTTTGCGGGCCCCCACAGTAGGTGCACGTGAGACTCAATGCCCGCCAGAATTCCAGG 115065
||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | |
 Cp   1309 TTTGAGCTTAAGCTCTCCACATGATGACAGTAAAGATGTTCTCTCTCAACTCT 1310
||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | |
 Db 115066 ATATGTTTTCCACATCTCCTTTGTAAGAACATTCACAGTCTTACCCCTGAGAGCTCTG 115125
||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | |
 Cp   1249 ATGTGATGTTCCACATCTCCTTTAAAAGAACTTCAACTGTCTTCCCTCTGAGGCTCTG 1190
||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | | ||||| . | |


RESULT      10              HSCU177E8    40822 bp       DNA                PRI           27-JUL-1996
LOCUS                                  Human DNA sequence from cosmid CU177E8, between markers DXS366 and
DEFINITION                                DXS87 on chromosome X.
ACCESSION                               Z68694
NID                                     g1160468
KEYWORDS                                 X.
SOURCE                                   human.
ORGANISM                                Homo sapiens
Eukaryote; mitochondria eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS                                Williamson,H. and Wilkinson,P.
TITLE                                  Direct Submission
JOURNAL                                      Submitted (16-JAN-1995) Sanger Centre, Hinxton, Cambridgeshire,
                                           CB10 1RQ, England. E-mail enquiries: humpdes@sanger.ac.uk
COMMENT
CU177E8. It may be shorter because we have inserted of clone
sections once, or longer because we arrange for a small overlap
between neighbouring submissions.
The true left end of clone CU177E8 is at 1 in this sequence. The
true right end of clone CU177E8 is at 40822.
The true right end of clone V693A8 is at 34707.
CU177E8 is from the human chromosome X-specific cosmid library.
location/Qualifiers
FEATURES
SOURCE
1..40822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="CU177E8"
2..82
/partial
/note="Alu repeat: matches 210..291 of consensus"
88..119
/note="8 copies of 4 mer 100 % conserved"
88..123
/note="6 copies of 6 mer 94 % conserved"
88..121
/note="17 copies of 2 mer 100 % conserved"
304...383
/note="1L element fragment"
1367..1420
/note="3 copies of 18 mer 91 % conserved"
1367..1414
/note="8 copies of 6 mer 100 % conserved"
1367..1414
/note="12 copies of 4 mer 100 % conserved"
1367..1414
/note="24 copies of 2 mer 100 % conserved"
4399...5185
/note="match: multiple ESTs"
6535..6564
repeat_region

```

```
repeat_region /note="15 copies of 2 mer 87 & conserved"
/note="7643
/note="L1 element fragment"
8303. .8594
repeat_region /partial
/note="Alu repeat: matches 308. .1 of consensus"
11666. .11956
/note="Alu repeat: matches 1. .306 of consensus"
12738. .12797
/note="L1 element fragment"
12800. .12847
/note="12 copies of 4 mer 81 & conserved"
12802. .12843
/note="7 copies of 6 mer 83 & conserved"
12858. .13167
/note="L1 element fragment"
13344. .13522
/note="L1 element fragment"
13615. .13840
/note="L1 element fragment"
13908. .13970
/note="L1 element fragment"
14029. .14166
/note="L1 element fragment"
14326. .14367
/note="L1 element fragment"
14431. .14481
/note="L1 element fragment"
14796. .15633
/note="L1 element fragment"
15713. .15972
/note="L1 element fragment"
16034. .16152
/note="L1 element fragment"
16193. .16672
/note="L1 element fragment"
17297. .17622
/note="MER42A element fragment"
17321. .17511
/note="MER42C element fragment"
18153. .18469
/note="MER42A element fragment"
19454. .19502
repeat_region /partial
/note="Alu repeat: matches 1. .51 of consensus"
19506. .19606
repeat_region /partial
/note="Alu repeat: matches 206. .308 of consensus"
21120. .21155
/note="6 copies of 6 mer 100 & conserved"
21120. .21157
/note="19 copies of 2 mer 100 & conserved"
21120. .21155
/note="9 copies of 4 mer 100 & conserved"
21710. .21939
/note="L1 element fragment"
22093. .22173
/note="L1 element fragment"
23772. .23873
/note="match: multiple ESTs"
24263. .24341
/note="match: multiple ESTs"
24926. .25954
/note="match: multiple ESTs"
28904. .28963
/note="match: multiple ESTs"
28909. .28962
/note="20 copies of 3 mer 85 & conserved"
28909. .28962
/note="3 copies of 18 mer 87 & conserved"
28924. .28965
/note="7 copies of 6 mer 95 & conserved"
29721. .29752
/note="16 copies of 2 mer 94 & conserved"
29721. .29752
repeat_region
```

```
repeat_region /note="8 copies of 4 mer 94 & conserved"
29723. .29752
/note="5 copies of 6 mer 93 & conserved"
30954. .31012
repeat_region /note="L1 element fragment"
31180. .31296
repeat_region /note="L1 element fragment"
31336. .31525
repeat_region /note="L1 element fragment"
33419. .33503
/note="MLTID element fragment"
33616. .33814
/note="MLTID element fragment"
34325. .34599
/note="Alu repeat: matches 1. .308 of consensus"
34646. .34937
/note="Alu repeat: matches 1. .308 of consensus"
34938. .34961
/note="6 copies of 4 mer 100 & conserved"
35011. .35050
/note="20 copies of 2 mer 100 & conserved"
35011. .35050
/note="10 copies of 4 mer 100 & conserved"
35011. .35052
/note="7 copies of 6 mer 98 & conserved"
37252. .37323
/note="24 copies of 3 mer 82 & conserved"
37252. .37323
/note="12 copies of 6 mer 82 & conserved"
37252. .37323
/note="4 copies of 18 mer 82 & conserved"
37866. .38153
/note="Alu repeat: matches 1. .308 of consensus"
39032. .39079
/note="12 copies of 4 mer 94 & conserved"
39038. .39077
/note="20 copies of 2 mer 100 & conserved"
39038. .39079
/note="7 copies of 6 mer 98 & conserved"
39830. .40086
/note="L1 element fragment"
40279. .40702
/note="L1 element fragment"
repeat_region /note="L1 element fragment"
40279. .40702
BASE COUNT 13043 a 7504 c 8664 g 11611 t
ORIGIN
Query Match 4.4%; Score 81; DB 22; Length 40822;
Best Local Similarity 71.4%; Pred. No. 3,62e-43;
Matches 152; Conservative 0; Mismatches 59; Indels 2; Gaps 2;
Db 19916 CATACATTATGTCAGGGAATGTGTGAC-GGAG-GATGCGAGAAAGATTCTTCTTCT 19973
|| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 789 CACACATCATGCGAGGAATGTGTGAACAGAGAGATGTGAAGAGCCTCTTCT 848
Db 19974 GGACACAAACATCTAAGTAGATATAAGCCAGGCTCTGTGTACTACTAGTGGCCAAA 20033
|| || || || || || || || || || || || || || || || || || || || || || || ||
Qy 849 GGGGCTGACATCAACAAAGGGGAATGGCCAGGCTCTGTGTACTACTCGGAAGAAA 908
Db 20034 ACGTTGGTGACATCGCTGTTGTGCCAAGTGATGATGATGCTGAATGCTTTCAT 20093
|| ||||| || || || || || || || || || || || || || || || || || || || || ||
Qy 909 ACTGAGTGGGGTATCCACAGCTCAGCCAGCTGATGAGATGTGACATGCTGCTCATGCC 968
Db 20094 CTGAAGGGGCAATCATCTCAGATGAGGCTGC 20126
|| || || || || || || || || || || || || || || || || || || || || || || ||
Qy 969 CTGAAGGCCACATCATTTCAAGATGAGGATTGC 1001
RESULT 11
LOCUS 166494 7218 bp DNA PAT 23-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
NID 92724471
KEYWORDS
```

[illegible]

```

Query Match          4.0%   Score73 : DB 22: Length 207;
Best Local Similarity 72.1%; Pred.No. 2,16e-36;
Matches    119; Conservative    0; Mismatches    46; Indels    0; Gaps    0;

Db      28  TTTAGACCTGGAGCAGATTTTGTTCATGCTCGGAGGAATGTTTTCGGCTCATACGAGTG 87
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1028 TTTTGGGGCCAGAGACTGCATCTCGTGTACTCTGGTGCGATTGGCTGGGCACAOTGAGTGC 1087

Db      88  TGCTGGANAAGTGTTTGAGGAGAACGACGCAAGCTCAAGCTCTTCTACGGGATGAGCTC 147
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1088 AGGTGTGAGGTCATCACGAGAGGAGGCCAAGAATCAAGCTCTTCATGAAATGAGTTC 1147

Db      148  TGACACCCGCATGAACAAGCAGCAGCGAGGAGTTGCTTAATACAG 192
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      1148 TGAATGGCCATGAAGAAGTATGCTGGGGGGCTGTGCTAGTACAG 1192


RESULT    13
LOCUS     ECOAMPISM             4791 bp       DNA              BCT               12-JAN-1995
DEFINITION Escherichia coli ampD gene: guinolinate phosphoribosyltransferase
            (nadc) gene: prepilin-like peptidase dependent protein (ppdD) gene:
            hopB, hopc genes; GMP reductase (guac) gene.
ACCESSION L28105
NID       9456039
KEYWORDS  GMP reductase; ampD gene; guac gene; hopB gene; hopc gene; nadc
            gene; ppdD gene; prepilin-like peptidase dependent protein;
            guinolinate phosphoribosyltransferase.
SOURCE    Escherichia coli (strain K-12) DNA.
ORGANISM  Escherichia coli
           Escherichia: Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE 1 (sites)
AUTHORS   Andrews,S.C. and Guest,J.R.
TITLE     Nucleotide sequence of the gene encoding the GMP reductase of
           Escherichia coli K12
JOURNAL   Biochem. J. 255 (1), 35-43 (1988)
MEDLINE   89061679
REFERENCE 2 (sites)
AUTHORS   Lindquist,S., Galleni,M., Lindberg,F. and Normark,S.
TITLE     Signalling proteins in enterobacterial Ampc beta-lactamase
           regulation
JOURNAL   Mol. Microbiol. 3 (8), 1091-1102 (1989)
MEDLINE   90113890
REFERENCE 3 (sites)
AUTHORS   Hobbs,M. and Mattick,J.S.
TITLE     Common components in the assembly of type 4 fimbriae. DNA transfer
           systems, filamentous phage and protein-secretion apparatus: a
           general system for the formation of surface-associated protein
           complexes
JOURNAL   Mol. Microbiol. 10, 233-243 (1993)
MEDLINE   95020523
REFERENCE 4 (bases 1 to 4791)
AUTHORS   Whitchurch,C.B. and Mattick,J.S.
TITLE     Escherichia coli contains a set of genes homologous to those
           involved in protein secretion. DNA uptake and the assembly of
           type-4 fimbriae in other bacteria
JOURNAL   Gene 150 (1), 9-15 (1994)
MEDLINE   95047556
FEATURES
         source          Location/Qualifiers
                     1..4791
                        /organism="Escherichia coli"
                        /strain="K-12"
                        /db_xref="taxon:562"
                        complement(1..290)
                        /gene="ampD"
                        complement(<1..290)
                        /gene="ampD"
                        /citation=[2]
                        /codon_start=1
                        /db_xref="PID:g456040"
                        /translation="MLEEGQWLVGARRVSPHYDCRDPDETPTLLVVYHNSLPGEEFG

```

BASE COUNT	1150 a	1276 c	1257 g	1108 t
ORIGIN	<pre> /citation=[1] /codon_start=3 /product="GMP reductase" /db_xref="PID:g456045" /transl_table=11 /translation="MAGHESSGRIVYENGKFMLEFGMSSESAMRRHYGVAYEYAA AEGKYALPLRGVENVETARDILGLNSACTYVGSRLKELRTTFIVQDEENRIFNM NL" </pre>			
Query Match	3 8%	Score 70;	DB 13;	Length 4791;
Best Local Similarity	61.4%;	Pred. No. 6,99e-34;		
Matches 188;	Conservative 0;	Mismatches 118;	Indels 0;	Gaps 0;
Db 4486	TGAAGATCGGGTTTCTCTGTTCTCTGCACACCAATAAAGTGGTGGTCACTCTT	4545		
Cp 1368	TGAAATATGGATTACACCTCGCTGGGTGACTCGGATGAGAGTATGTTCTCTGCTCAACTTT	1309		
Db 4546	TCAGGCGGTGAAGCCCAAGTATGTACAAGCTGAACGACGAGCCGCCAAAATATCTCGCG	4605		
Cp 1308	TGAGGTATAGTGTCTGCCACATAGGACACGTAGAAGCGGATCCCTCTGATGATGTCTCGGA	1249		
Db 4606	CGGATTTTCAACCGGGGCTTCGACAGCGGCACCTTAACGCTTTTACCTTCTGCTGGCGAT	4665		
Cp 1248	TGGATATGTTCCACATCTCTTTAAAGGAACTTCCACTGTCTTCCCTGAGGCTCTGT	1189		
Db 4666	ATTCGCGAAGCCCGGCAACGTGAGCTTTCATCGGAGACTCGGAGCTCATGGCGTAGAACA	4725		
Cp 1188	ACTACGACCAACCCCGCCACGATACCTTCTTCAATGAGGCATTTTCAGAACTCATTCATAGAAAGA	1129		
Db 4726	GCATTAATTTCTCGCCGCTTCTCTCAACGATGACGACGCGCCCTCTCTTCGTGGCCCGCA	4785		
Cp 1128	GCTGTACTTCTTGGCAATCCCTCTCGATGAGCTCACCAACCTGACTCATCTGTGCCAGCA	1069		
Db 4786	GCATGC 4791			
Cp 1068	GCATGC 1063			
RESULT 14				
LOCUS	S73035S2	159 bp	DNA	PRI 10-JUL-1992
DEFINITION	guanosine monophosphate reductase [human, genomic, 159 nt, segment 2 of 9].			
ACCESSION	S73043			
NID	g241125			
KEYWORDS	2 of 9			
SEGMENT	human.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryotes: mitochondrial eukaryotes; Metazoa: Chordata; Vertebrata: Eutheria; Primates; Catarrhini: Hominoidea; Homo. 1 (bases 1 to 159)			
AUTHORS	Kondo,T., Kanno,H., Chang,L. and Yoshida,A.			
TITLE	Genomic structure and expression of human guanosine monophosphate reductase			
JOURNAL	Hum. Genet. 88 (2), 219-224 (1991)			
MEDLINE	92098099			
REMARK	GenBank staff at the National Library of Medicine created this entry [NCBI gidsq 73043] from the original journal article. This sequence comes from Figure 5.			
FEATURES	Location/Qualifiers			
SOURCE	1..159			
BASE COUNT	35 a	39 c	41 g	44 t
ORIGIN	<pre> /organism="Homo sapiens" /db_xref="taxon:9606" </pre>			
Query Match	3 6%	Score 66;	DB 22;	Length 159;
Best Local Similarity	81.1%;	Pred. No. 1.44e-30;		
Matches 86;	Conservative 0;	Mismatches 20;	Indels 0;	Gaps 0;
Db 26	AGGTGATCTTGAACGCACTTCAGTTGCAAAATTCAAAGACAGACCTACAGGATTC	85		

This Page Blank (uspto)


```
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KM Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1993; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI: 93-378844/8.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3, Page 14; 23pp. English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 2.4%; Score 44; DB 9; Length 91;
Best Local Similarity 0.0%; Pred. No. 4,42e-13;
Matches 0; Conservative 44; Mismatches 0; Indels 0; Gaps 0;

Db 13 vhsyvvvvhshhshvhhvhhvshvvvvhvhhvhhvhhv 56
CP 1036 GCCCAAAAGCCTTGCCACATCCCGACAGACGTCGCAACCTCC 993

RESULT 3
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KM E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /*tag= b

EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Kotivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.4%; Score 45; DB 1; Length 204;
Best Local Similarity 8.7%; Pred. No. 9.73e-14;
Matches 8; Conservative 56; Mismatches 27; Indels 1; Gaps 1;
```

```
Db 93 hyrmnbvnyrdynrdaawcyrrsvkydcynachdhvypbbvnyvnhnncn 152
CP 101 CTGGAAAGGCTGTGACAGACCTGACGGTGCAAT-CTTACCCCGCTACACTTCCTTC 43

Db 153 ccdbnhvchvbnhbnhrwayvrhdardv 184
CP 42 TTTCAACCCCTGCAGAGCATGGGCGCAAGG 11

RESULT 4
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KM E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69
FT /tag= a
FT /function= multiple cloning site
FT primer_bind 187..204
FT /*tag= b

EP-285123-A.
PD 05-MAY-1988.
PF 30-MAR-1988; 105163.
PR 03-APR-1987; US-034819.
PA (SUSO) SUOMEN SOKERI OY.
PI Lehtovaara P, Knowles J, Kotivula A, Bamford J, Reinikainen T;
DR WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PS Disclosure; p; English.
CC Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.
SQ Sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

Query Match 2.3%; Score 42; DB 1; Length 204;
Best Local Similarity 10.2%; Pred. No. 8.90e-12;
Matches 11; Conservative 57; Mismatches 40; Indels 0; Gaps 0;

Db 78 dchvycgymttthhyrmnbvnyrdynrdaawcyrrsvkydcynachdhv 137
QY 1270 TCCGCTACACTGTACATGATGGAGACGTAAGCTCAAGAAATTGACGAGAACTA 1329

Db 138 ybbvnyvnhnncnccbnhchvbnhbnhrwayvrhdardv 185
QY 1330 CCTTCATCCGAGTACCCAGCAGAGTGAATCAATCTTCAGTAGGCGCT 1377

RESULT 5
ID Q06261 standard; DNA; 1829 BP.
AC Q06261;
DT 28-JAN-1991 (first entry)
DE Sequence encoding IMP dehydrogenase gene.
KW Inosine; guanosine; 5' inosinic acid; 5' guanylic acid; ds.
OS bacillus sp.
PN EP-393969-A.
PD 24-OCT-1990.
PF 17-APR-1990; 304080.
PR 19-APR-1989; JP-101084.
PA (TAKE ) TAKEDA CHEMICAL IND KK.
```

ID	TSAR	Location/Qualifiers	Score	DB	Length	Indels	Gaps
PI	Myagawa K, Kanazaki N, Hasegawa T;						
DR	WPI: 50-322440/43.						
PT	Introduction of different expressible promoter into DNA - used						
PT	for improved prodn. of inosine and/or guanosine						
PS	Disclosure: Fig 2: 21pp: English.						
CC	Sequence may be incorporated into a plasmid vector under the control						
CC	of a foreign expressible promoter, and used to transform a Bacillus						
CC	bacteria. New promoter may be used to express the gene at higher or						
CC	lower levels in pure form.						
CC	The products inosine and/or guanosine are the raw materials for the						
CC	production of 5'-IMP and 5'-GMP, flavour components of						
CC	"katsunobushi" and "shiitake mushroom".						
CC	Sequence 1829 BP; 562 A; 359 C; 420 G; 488 T;						
SO							
Db	Query Match	2.3%; Score 43; DB 1; Length 1829;					
Db	Best Local Similarity 63.5%; Pred. No. 1.99e-12;						
Db	Matches 101; Conservative 0; Mismatches 58; Indels 0; Gaps 0;						
Qy	1013 cattattgctggaacgctgacacagctgaagcgacaaagcgcttatcgaaagtgcg 1072						
Qy	794 CATCATGCGACAGGATGCTGTACACGAGACAGATGCTAGAAAGCCTCATCTTCTGGGCG 853						
Db	1073 agacgtgtcaaaattggaataggagccgctgtcaattgtactacacagttgttgaagcgg 1132						
Qy	854 TGACATCATCAAAAGGGGAATTGGGCCACGCGCTCTGTGTACTACTGGAAGAAACTGG 913						
Db	1133 ggtggtgtctccgcaaatataacgaattatgattgtgc 1171						
Qy	914 AGTGGGCTATTCACACGCTCACGCCGACGATGATGATGTGC 952						
RESULT	6						
ID	Q70469 standard; DNA: 114 BP.						
AC	Q70469:						
DT	07-APR-1995 (first entry)						
DE	Generic DNA sequence to generate a random TSAR peptide library.						
KW	TSAR: totally synthetic affinity reagent; synthetic; binding domain;						
KW	effector domain; concatenated heterofunctional protein; linker;						
KW	direct; rapid; detection; screening; treatment; generic; ss.						
OS	Synthetic.						
FT	Key	Location/Qualifiers					
FT	misc_feature	55..60					
FT		/note= "a					
FT		sequence of 6,9 or 12 nucleotides (see					
FT		comments)"					
PN	W09418318-A.						
PD	18-AUG-1994.						
PF	01-FEB-1994; U00977.						
PR	01-FEB-1993; US-013416.						
PR	30-DEC-1993; US-176500.						
PR	31-JAN-1994; US-189331.						
PA	(UYNC-) UNIT NORTH CAROLINA.						
PI	Fowlkes DM, Kay BK;						
DR	WPI: 94-219739/34.						
PT	Identifying proteins or peptide(s) which bind a ligand - by						
PT	screening a recombinant vector library expressing fusion proteins						
PT	comprising a binding domain and an effector domain						
PS	Disclosure: Page 35: 25pp: English.						
CC	Q70469 is a generic DNA sequence used to generate random TSAR peptide						
CC	library. The generic formula can be represented as follows: X(TGC)(NNB)10-						
CC	(TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC). X and Y are flanking restriction						
CC	sites (X is not the same as Y) that are not specified further. This						
CC	sequence generates peptides that are cloverleaf in structure. Other						
CC	generic sequences are shown in Q70465-68. Other specific peptides						
CC	generated by these generic sequences are shown in RS5150-54. TSARs are						
CC	concatenated heterofunctional proteins or peptides, comprising at least						
CC	two functional regions - a binding domain with affinity for a ligand and						
CC	a second effector peptide portion that is chemically or biologically						
CC	active. They may further comprise a linker peptide between the 2 domains						
CC	The oligonucleotides are also designed so that the expressed peptide						
CC	contains 2 or 4 cysteine residues positioned in, or flanking, the						
CC	unpredicted or variant residues. These residues confer some degree of						

[illegible]

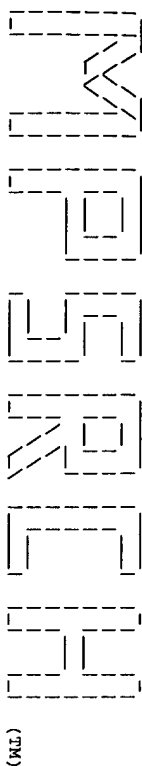
CC	production. The TSARs are easily characterised and have designed
CC	activity allowing direct and rapid detection in a screening process.
SO	Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
Query Match	1.88; Score 34; DB 12; Length 114;
Best Local Similarity	2.78; Prod. No. 9.63e-07;
Matches	3; Conservative 34; Mismatches 75; Indels 0; Gaps 0;
Db	3 bnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnnnnnn 62
Oy	476 GGTCCCATCATGTCGCAATATGATAGTGGGACCTTGATGACCAAGTTCT 535
Db	63 bnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnb 114
Oy	536 CCGTAGATTCTCTCTCTCAGTCGTCCTCATTAAGACATATAGCCTCCTTGG 587
RESULT	8
ID	070467 standard; DNA; 114 BP.
AC	070467.
DE	05-APR-1995 (first entry)
DT	Generic DNA sequence to generate a random TSAR peptide library.
KS	TSAR: totally synthetic affinity reagent; synthetic; binding domain;
KW	effector domain; concatenated heterofunctional protein; linker;
KM	direct; rapid; detection; screening; treatment; generic; ss.
OS	Synthetic.
FS	Key
FT	Location/Qualifiers
FT	misc_feature
FT	55..60
FT	/tag= a
FT	/note= "this sequence represents 'Z'; Z can be a
FT	sequence of 6, 9 or 12 nucleotides (see
FT	comments)"
PN	W09418318-A.
PD	18-AUG-1994.
PR	01-FEB-1994; U00977.
PR	01-FEB-1993; US-013416.
PR	30-DEC-1993; US-176500.
PR	31-JAN-1994; US-189331.
PA	(UYNC-) UNIV NORTH CAROLINA.
PI	Fowlkes DM, Kay BK;
DR	WPI: 94-279739/34.
DR	P-PSDB: R65153.
PT	Identifying proteins or peptide(s) which bind a ligand - by
PT	screening a recombinant vector library expressing fusion proteins
PS	comprising a binding domain and an effector domain
PS	disclosure; Page 35; 255PP; English.
CC	070467 is a generic DNA sequence used to generate random TSAR (Totally
CC	Synthetic Affinity Reagents) peptides. This generic formula can also be
CC	represented as follows: X(NNB)16(TGC)(NND)16(TTC)(NNB)15. X
CC	and Y are flanking restriction sites (X is not the same as Y) that are
CC	not specified further. Other generic sequences are shown in 070466-68.
CC	Other specific peptides generated by these generic sequences are shown in
CC	R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC	comprising at least two functional regions - a binding domain with
CC	affinity for a ligand and a second effector peptide portion that is
CC	chemically or biologically active. They may further comprise a linker
CC	peptide between the 2 domains. The oligonucleotides are also designed so
CC	that the expressed peptide contains 2 or 4 cysteine residues positioned
CC	in, or flanking, the unpredicted or variant residues. These residues
CC	confer some degree of conformational rigidity to the peptides. The TSARs
CC	or compens, comprising a TSAR binding domain can be used in vivo to
CC	deliver a chemically or biologically active moiety, eg. metal ion,
CC	radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC	cell. They can also replace the function of macromolecules, eg.
CC	monoclonal or polyclonal antibodies and therefore circumvent the need for
CC	complex methods of hybridoma formation or in vivo antibody production.
CC	The TSARs are easily characterised and have designed actively allowing
CC	direct and rapid detection in a screening process.
SO	Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
Query Match	1.88; Score 34; DB 12; Length 114;
Best Local Similarity	2.78; Prod. No. 9.63e-07;
Matches	3; Conservative 34; Mismatches 75; Indels 0; Gaps 0;
Db	3 bnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnbhnb 114
Oy	536 CCGTAGATTCTCTCTCTCAGTCGTCCTCATTAAGACATATAGCCTCCTTGG 587

[illegible]

AC	Q70465:
DT	05-APR-1995 (first entry)
DE	Generic DNA sequence to generate a random TSAR peptide library.
KW	TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KM	effector domain; concatenated heterofunctional protein; linker;
KW	direct; rapid; detection; screening; treatment; generic; ss.
OS	Synthetic.
FH	Key
FT	misc_feature
FT	55..60
FT	/tag- a
FT	/note- "this sequence represents '%'; % can be a
FT	sequence of 6, 9 or 12 nucleotides (see
FT	comments)"
PN	
PT	MO9418318-A.
PD	18-AUG-1994.
PF	01-FEB-1994; U00977.
PR	01-FEB-1993; US-013416.
PR	30-DEC-1993; US-176500.
PR	31-JAN-1994; US-189331.
PA	(HYNC) UNIV NORTH CAROLINA.
PI	Fowlkes DM, Kay BK.
DR	WPI; 94-279739/34.
DR	P-PDB; R65150 and R65151.
PT	Identifying proteins or peptide(s) which bind a ligand - by
PT	screening a recombinant vector library expressing fusion proteins
PT	comprising a binding domain and an effector domain
PS	Disclosure; page 35; 25pp; English.
CC	Q70465 is a generic DNA sequence used to generate random TSAR (Totally
CC	Synthetic Affinity Reagents) peptides.This generic formula can also be
CC	represented as follows: X(NNB)6(TGC)(NNB)11(TGC)(NNB)3Y. X
CC	and Y are flanking restriction sites (X is not the same as y) that are
CC	not specified further. Other generic sequences are shown in Q70466-68.
CC	Other specific peptides generated by these generic sequences are shown in
CC	R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC	comprising at least two functional regions - a binding domain with
CC	affinity for a ligand and a second effector peptide portion that is
CC	chemically or biologically active.They may further comprise a linker
CC	peptide between the 2 domains.The oligonucleotides are also designed so
CC	that the expressed peptide contains 2 or 4 cysteine residues positioned
CC	in, or flanking, the unpredicted or variant residues. These residues
CC	confer some degree of conformational rigidity to the peptides. The TSARs
CC	or compms. comprising a TSAR binding domain can be used in vivo to
CC	deliver a chemically or biologically active moiety, eg. metal ion,
CC	radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC	cell. They can also replace the function of macromolecules, eg.
CC	monoclonal or polyclonal antibodies and therefore circumvent the need
CC	for complex methods of hybridoma formation or in vivo antibody
CC	production. The TSARs are easily characterised and have designed
CC	activity allowing direct and rapid detection in a screening process.
CC	Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
Query Match	1.8%; Score 34; DB 12; Length 114;
Best Local Similarity	3.6%; Pred.No. 9.63e-07;
Matches	4; Conservative 33; Mismatches 75; Indels 0; Gaps 0;
Dn	3 bnnbnbnbnbnbnbtgcnnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbnnnnnnn 62
Cp	903 TCCGAGTAGTACACACAGAGCGCTGGCCCAATTGCCACTTGATGTACAGCCCCAGAAA 844
Dn	63 bnnbn 114
Cp	843 GGATGACCTCTTCACCATCTCCTGTTACCAACATTCCTGCATGATGGT 792
RESULT	11
ID	Q70468 standard; DNA; 114 BP.
AC	Q70468;
DT	05-APR-1995 (first entry)
DE	Generic DNA sequence to generate a random TSAR peptide library.
KW	TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KM	effector domain; concatenated heterofunctional protein; linker;
KW	direct; rapid; detection; screening; treatment; generic; ss.
OS	Synthetic.

FH	Key	Location/Qualifiers
FT	misc_feature	55..60
FT		/tag= a
FT		/note= "this sequence represents 'Z'; Z can be a
FT		sequence of 6, 9 or 12 nucleotides (see
FT		comments)"
PN		
PD	M09A18318-A.	
PD	18-APR-1994.	
PF	01-FEB-1994; U00977.	
PR	01-FEB-1993; US-013416.	
PR	30-DEC-1993; US-176500.	
PR	31-JAN-1994; US-189331.	
PA	(UYNC-) UNIT NORTH CAROLINA.	
PI	Fowlkes DM, Kay BK;	
DR	WPI; 94-279739/34.	
DR	P-PDB; R65154.	
PT	Identifying proteins or peptide(s) which bind a ligand - by	
PT	screening a recombinant vector library expressing fusion proteins	
PT	comprising a binding domain and an effector domain	
PS	Disclosure; Page 35; 255pp; English.	
CC	Q70468 is a generic DNA sequence used to generate random TSAR (Totally	
CC	Synthetic Affinity Reagents) peptides. This generic formula can also be	
CC	represented as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(GC)(NNB)10Y. X	
CC	and Y are flanking restriction sites (X is not the same as Y) that are	
CC	not specified further. Other generic sequences are shown in Q70466-68.	
CC	Other specific peptides generated by these generic sequences are shown in	
CC	R65151-54. TSARs are concatenated heterofunctional proteins or peptides,	
CC	comprising at least two functional regions - a binding domain with	
CC	affinity for a ligand and a second effector peptide portion that is	
CC	chemically or biologically active. They may further comprise a linker	
CC	peptide between the 2 domains. The oligonucleotides are also designed so	
CC	that the expressed peptide contains 2 or 4 cysteine residues positioned	
CC	in, or flanking, the unpredicted or variant residues. These residues	
CC	confer some degree of conformational rigidity to the peptides. The TSARS	
CC	or compns. comprising a TSAR binding domain can be used in vivo to	
CC	deliver a chemically or biologically active moiety, eg. metal ion,	
CC	radioisotope, peptide, toxin or enzyme, to the specific target or on the	
CC	cell. They can also replace the function of macromolecules, eg.	
CC	monoclonal or polyclonal antibodies and therefore circumvent the need	
CC	for complex methods of hybridoma formation or in vivo antibody	
CC	production. The TSARs are easily characterised and have designed activity	
CC	allowing direct and rapid detection in a screening process.	
CC	Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;	
SO		
Query Match	1.8%; Score 33; DB 12; Length 114;	
Best Local Similarity	2.8%; Pred. No. 3, 88e-06;	
Matches	3; Conservative 33; Mismatches 73; Indels 0; Gaps 0;	
Df	3 bmbnbmbnbnbnbnbnbnbnbnbnbtgcmbnbnbnbnbnbnbnbnbnnnnnn 62	
Cp	1652 GGATGTTGTGTGGTCCAGATTCCATTGATTTCTTCATCATTTCTCGACTG 1593	
Df	63 bmbnbmbnbnbnbnbnbtgcnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbn 111	
Cp	1592 TTTCCTTCCTTGCTTCCTCCAGTAGTGCTTGCGGACATTTGTGTGT 1544	
RESULT	12	
ID	Q70467 standard; DNA; 114 BP.	
AC	Q70467;	
DT	05-APR-1995 (first entry)	
DE	Generic DNA sequence to generate a random TSAR peptide library.	
KM	TSAR: totally synthetic affinity reagent; synthetic; binding domain;	
KW	effector domain: concatenated heterofunctional protein; linker;	
OS	direct; rapid; detection; screening; treatment; generic; ss.	
FN	Synthetic.	
FT	Key	Location/Qualifiers
FT	misc_feature	55..60
FT		/tag= a
FT		/note= "this sequence represents 'Z'; Z can be a
FT		sequence of 6, 9 or 12 nucleotides (see
FT		comments)"
PN		
PD	M09A18318-A.	
PD	18-APR-1994.	
PF	01-FEB-1994; U00977.	
PR	01-FEB-1993; US-013416.	
PR	30-DEC-1993; US-176500.	
PR	31-JAN-1994; US-189331.	
PA	(UYNC-) UNIT NORTH CAROLINA.	
PI	Fowlkes DM, Kay BK;	
DR	WPI; 94-279739/34.	
DR	P-PDB; R65154.	
PT	Identifying proteins or peptide(s) which bind a ligand - by	
PT	screening a recombinant vector library expressing fusion proteins	
PT	comprising a binding domain and an effector domain	
PS	Disclosure; Page 35; 255pp; English.	
CC	Q70468 is a generic DNA sequence used to generate random TSAR (Totally	
CC	Synthetic Affinity Reagents) peptides. This generic formula can also be	
CC	represented as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(GC)(NNB)10Y. X	
CC	and Y are flanking restriction sites (X is not the same as Y) that are	
CC	not specified further. Other generic sequences are shown in Q70466-68.	
CC	Other specific peptides generated by these generic sequences are shown in	
CC	R65151-54. TSARs are concatenated heterofunctional proteins or peptides,	
CC	comprising at least two functional regions - a binding domain with	
CC	affinity for a ligand and a second effector peptide portion that is	
CC	chemically or biologically active. They may further comprise a linker	
CC	peptide between the 2 domains. The oligonucleotides are also designed so	
CC	that the expressed peptide contains 2 or 4 cysteine residues positioned	
CC	in, or flanking, the unpredicted or variant residues. These residues	
CC	confer some degree of conformational rigidity to the peptides. The TSARS	
CC	or compns. comprising a TSAR binding domain can be used in vivo to	
CC	deliver a chemically or biologically active moiety, eg. metal ion,	
CC	radioisotope, peptide, toxin or enzyme, to the specific target or on the	
CC	cell. They can also replace the function of macromolecules, eg.	
CC	monoclonal or polyclonal antibodies and therefore circumvent the need	
CC	for complex methods of hybridoma formation or in vivo antibody	
CC	production. The TSARs are easily characterised and have designed activity	
CC	allowing direct and rapid detection in a screening process.	
CC	Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;	
SO		
Query Match	1.8%; Score 33; DB 12; Length 114;	
Best Local Similarity	2.8%; Pred. No. 3, 88e-06;	
Matches	3; Conservative 33; Mismatches 73; Indels 0; Gaps 0;	
Df	3 bmbnbmbnbnbnbnbnbnbnbnbnbtgcmbnbnbnbnbnbnbnbnbnnnnnn 62	
Cp	1652 GGATGTTGTGTGGTCCAGATTCCATTGATTTCTTCATCATTTCTCGACTG 1593	
Df	63 bmbnbmbnbnbnbnbnbtgcnbnbnbnbnbnbnbnbnbnbnbnbnbnbnbn 111	
Cp	1592 TTTCCTTCCTTGCTTCCTCCAGTAGTGCTTGCGGACATTTGTGTGT 1544	

This Page Blank (uspto)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPearch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Wed Jun 17 12:07:46 1998; MasPar time 1802.52 Seconds

Tabular output not generated. 1365.733 Million cell updates/sec

Title: >US-08-774-169-2
Description: (1-1843) from US08774169.seq
Perfect Score: 1843
N.A. Sequence: 1 CTAATACAGCCTCTGCC...AGAGAGCTATTGACTGT 1843
Comp: GATTTCGTGCGAGACGAGG...TCTCTCGAGTACGACGA

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 1759237 seqs, 667866413 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-est54

Database: genbank-est106

3:gb-est14 4:gb-est10 5:gb-est11 6:gb-est12 7:gb-est13
8:gb-est14 9:gb-est15 10:gb-est16 11:gb-est17
12:gb-est18 13:gb-est19 14:gb-est2 15:gb-est10
16:gb-est12 17:gb-est22 18:gb-est3 19:gb-est4 20:gb-est5
21:gb-est6 22:gb-est7 23:gb-est8 24:gb-est9 25:gb-est5
26:gb-est5

Statistics: Mean 11.691; Variance 1.920; scale 6.088

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	604	32.8	635	8	AA507892	nh75c02.s1 NCI_CGAP.Br	0.00e+00
2	568	30.8	646	22	W60937	zdz7e06.r1 Soares feta	0.00e+00
3	549	29.8	632	22	W65156	zcs8e05.r1 Soares para	0.00e+00
4	502	27.2	731	23	AA172963	ms58b04.r1 Life Tech m	0.00e+00
5	491	26.6	494	13	AA009421	ze82c05.r1 Soares feta	0.00e+00
6	489	26.5	501	19	N28887	yx63a11.r1 Homo sapien	0.00e+00
7	485	26.3	546	13	N30305	yx69h06.s1 Homo sapien	0.00e+00
8	464	25.2	472	13	AA778158	zif5b09.s1 Soares feta	0.00e+00
9	465	25.2	474	19	N35234	yy21d05.s1 Homo sapien	0.00e+00
10	465	25.2	478	7	AA456771	zw27f02.r1 Soares ovar	0.00e+00
11	460	25.0	599	9	AA524317	ng33h05.s1 NCI_CGAP.Co	0.00e+00
12	452	24.5	473	11	AA654123	ntl0b06.s1 NCI_CGAP.Pr	0.00e+00
13	450	24.4	463	19	H99254	yx21h07.s1 Homo sapien	0.00e+00

14	448	24.3	642	10	AA616194	vo96g03.r1 Soares mous	0.00e+00
15	445	24.1	474	16	AA830800	oc556d03.s1 NCI_CGAP_GC	0.00e+00
16	441	23.9	482	19	N39743	yx92h03.r1 Homo sapien	0.00e+00
17	435	23.6	472	22	W31386	zb95g03.s1 Soares para	0.00e+00
18	432	23.4	452	8	AA480954	aa28e03.r1 NCI_CGAP_GC	0.00e+00
19	432	23.4	483	22	W60872	zdz27e06.s1 Soares feta	0.00e+00
20	425	23.1	438	15	AA846516	a185h08.s1 Soares para	0.00e+00
21	416	22.6	432	11	AA683175	ag94d06.s1 Stratagene	0.00e+00
22	416	22.6	492	23	AA199147	nw42f08.r1 Soares mous	0.00e+00
23	415	22.5	428	23	AA009725	ze82c05.s1 Soares feta	0.00e+00
24	413	22.4	451	13	AA761349	nz23h03.s1 NCI_CGAP_GC	0.00e+00
25	402	21.8	522	5	AA004431	zn87a04.r1 Soares feta	0.00e+00
26	400	21.7	423	12	AA112533	zn28f07.r1 Stratagene	0.00e+00
27	394	21.4	446	4	AA306422	ESR177585 Jurkat T-cell	0.00e+00
28	394	21.4	464	12	N95282	zb60c12.s1 Soares feta	0.00e+00
29	394	21.4	464	12	AA113312	zn28f07.s1 Stratagene	0.00e+00
30	391	21.2	406	11	AA653970	nr09a05.s1 NCI_CGAP_Pr	0.00e+00
31	386	20.9	408	9	AA551600	nj57a05.s1 NCI_CGAP_Pr	0.00e+00
32	385	20.9	414	23	W86394	zh54d12.s1 Soares feta	0.00e+00
33	383	20.8	418	23	N55379	yy47h07.r1 Soares feta	0.00e+00
34	384	20.8	457	22	W15516	zb91a09.s1 Soares para	0.00e+00
35	379	20.6	406	20	N78792	z994d03.s1 Homo sapien	0.00e+00
36	376	20.4	378	19	N42029	yy69h06.r1 Homo sapien	0.00e+00
37	374	20.3	490	13	AA740276	ob27c01.s1 NCI_CGAP_K1	0.00e+00
38	373	20.2	500	3	T48085	yb24g10.s1 Homo sapien	0.00e+00
39	369	20.0	404	16	AA181712	zp49b02.s1 Stratagene	0.00e+00
40	368	20.0	410	19	N44897	yy21d05.r1 Homo sapien	0.00e+00
41	362	19.6	385	13	AA730615	ny41h05.s1 NCI_CGAP_Bw	0.00e+00
42	360	19.5	374	12	AA699746	zj77g10.s1 Soares feta	0.00e+00
43	357	19.4	386	22	H94052	yy15f02.s1 Soares feta	0.00e+00
44	348	18.9	355	4	AA342623	EST48204 Fetal spleen	0.00e+00
45	344	18.7	432	16	AA870746	vq26a09.r1 Barstead st	0.00e+00

ALIGNMENTS

RESULT 1
LOCUS AA507892 635 bp mRNA
DEFINITION nh75c02.s1 NCI_CGAP.Br.1 Homo sapiens cDNA clone IMAGE:964322
similar to gb:M24470 GMP REDUCTASE (HUMAN);, mRNA sequence.

ACCESSION AA507892
NID 92244331
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae;
Homo.

REFERENCE 1 (bases 1 to 635)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 1257 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 448.
Location/Qualifiers

FEATURES

1..635
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified

OY	1393	ATCCGAGCACCCACGACGTGAATCATCTTCACTAGTAGGGTGCTAACCTGCACCA-GT	1393
Db	302	TCTACCTGCCAAGCACCACAAGTAATTACTTACCATGGGGCATCCCAGTTGGGTCCTCACC	361
OY	1394	TCTACCTGCCAAGCACCACA-GTACTTACCATGGGGCATCCCCAAGT-GGGGTCCTCACC	1451
Db	362	CATCCACGCTACGAGCTCGTATTAATTGTCATTTCCCTGGTCTGCACCTCGAAGG	421
OY	1452	CATCCACGCTACGAGCTCGTATTAATTGTCATTTCCCTGGTCTGCACCTCGAAGG	1511
Db	422	CTCTCGAGTAAGTCTGTACTTCTTATCTGCACACACAAAATGCCCAAGCATCAGTG	481
OY	1512	CTCTCGAGTAAGTCTGTACTTCTTATCTGCACACACAAAATGCCCAAGCATCAGTG	1571
Db	482	GCGAGAAGCAAGAAAGCAACAGTCTGAGAAAAATGATGCAAGAAA-TCAAATGGGAATC	540
OY	1572	GCGAGAAGCAAGAAAGCAACAGTCTGAGAAAAATGATGCAAGAAAATCAAATGGGAATC	1631
Db	541	TGGGGACCCA-CACAACTCTGTGAAGATTATTTNNAAGAAA-GATGCTGATTG-TACATA	597
OY	1632	TGGGGACCCAACAACTCTGTGAAGATTATTTNNAAGAAAAGATGCTGATTGTCATATA	1691
Db	598	A-TCCCTTACATGGGCT-GGCTAGAG-AAGGAGGCGNTTGA-TCATGTTTT	646
OY	1692	AATCTTTACATGGGCTTGCTTACAGGAGCGAGGCTTTTGAATCATGTTTT	1744

RESULT	3	W56156	632 bp	mRNA	EST	11-OCT-1996
LOCUS	Z6588e5.r1					
DEFINITION	328528 5', mRNA sequence.					
ACCESSION	W56156					
NID	g1358045					
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 632)					
AUTHORS	Hallier,L., Clark,N., Dubouque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.					
TITLE	The WashU-Merck EST Project					
JOURNAL	Unpublished (1995)					
COMMENT	Contact: Wilson RK WashU-Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watsn.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1472 Std Error: 0.00 Seq primer: mod.BEBA+ET High quality sequence stop: 404. Location/Qualifiers 1..632 /organism="Homo sapiens" /note="Organ: parathyroid gland; Vector: pUT3D (Pharmacia) With a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'- TGTATCAATCTGAAAGGGAGGGCCGACCAATTTTTTTTTTTTTTTTTT T-3'] , double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pUT33 vector (Pharmacia). Library went through one round of					

Query Match	Best Local Similarity	95.7%	Pred. No. 0.00e+00;	Mismatches	0;	Indels	7;	Gaps	7;
Matches	605;	Conservative							
Db	1	CTCTCTTTATTGAGCTATTTC	CAATATACAAACACTTCAGAGCTTTT	TAGATATGTAAG	60				
Cp	1829	CTCTCTTTATTGAGCTATTTC	CAATATACAAACACTTCAGAGCTTTT	TAGATATGTAAG	1770				
Db	61	GTCGAATTTAGTGAAGGGATTA	CAACAATATGATTTAAAGCGCCCTCTGACC	120					
Cp	1769	GTCGAATTTAGTGAAGGGATTA	CAACAATATGATTTAAAGCGCCCTCTGACC	1710					
Db	121	AAGGCATGTAAAGATTTATGT	ACCAATGACATCTTTCTTTTAAATCTTCAGGA	180					
Cp	1709	AAGGCATGTAAAGATTTATGT	ACCAATGACATCTTTCTTTTAAATCTTCAGGA	1650					
Db	181	TGTTGTGTGGGTGCCCAAGT	TCCCATTTGATTTCTTGATCATATTTCTCAAGCTGTT	240					
Cp	1649	TGTTGTGTGGGTGCCCAAGT	TCCCATTTGATTTCTTGATCATATTTCTCAAGCTGTT	1590					
Db	241	GCTTCCTGCTCTCTCCCAAGT	AGTGCCTTGATTTGTTGTGTGACATATGAGAAGT	300					
Cp	1589	GCTTCCTGCTCTCTCCCAAGT	AGTGCCTTGATTTGTTGTGTGACATATGAGAAGT	1530					
Db	301	ACAGAGTTTACTGACAGAGACC	CTCGAGAGTGAACAACAGAGAATGACAAAGTAATACAGA	360					
Cp	1529	ACAGAGTTTACTGACAGAGACC	CTCGAGAGTGAACAACAGAGAATGACAAAGTAATACAGA	1470					
Db	361	GCTCAGTAGCTGGGATGGGT	GAGGACCCCACTTGGGATGCCCATGTAGAGTACTGGT	420					
Cp	1469	GCTCAGTAGCTGGGATGGGT	GAGGACCCCACTTGGGATGCCCATGTAGAGTACTGGT	1410					
Db	421	GCTTTGGGAGGGGTGAGAAC	TCTCAGGGTCTAGCAGCCCTCAGTGAAGATTGGATTAC	480					
Cp	1409	GCTTTGGGAGGGGTGAGAAC	TCTCAGGGTCTAGCAGCCCTCAGTGAAGATTGGATTAC	1352					
Db	481	TGCGGGGGGAGACTCCGGAT	TGAAGTAGTTCCTGCTCAAAATCCCTTGAGTTAGTGGCT	540					
Cp	1351	TGCTGGGTGACTC-GGAT	TGAAGTAGTTCCTGCTCAAAATCCCTTGAGTTAGTGGCT	1295					
Db	541	CCCAACATAGGTACACGTAN	AGCGGATTCGCGCAAGGATGTCTCCGATGGAGTGTCC	600					
Cp	1294	CCCAACATAGGTACACGTAN	AGCGGATTCGCGCTGCTA-GGATGTGTC-GGATGTGATGTTCCA	1237					
Db	601	CANCTCCCTTNAAGGAC	TTCCCTTCTCTT 632						
Cp	1236	CANCTCCCTTNAAGGAC	TTCCCTTCTCTT 1205						
RESULT	4	AA172963	731 bp	mrna	EST	23-DEC-1996			
LOCUS		ms58b04.r1	Life Tech mouse embryo 13 5dpc	10666014	Mus musculus				
DEFINITION		CDNA clone 615727 5'	similar to gb:M24470 GMP REDUCTASE (HUMAN);						
ACCESSION		AA172963							
NID		91752032							
KEYWORDS		EST.							
SOURCE		house mouse.							

REFERENCE	AUTHORS	TITLE JOURNAL COMMENT
Mus musculus	Eukaryote; mitochondrial eukaryotes; Metazoa; Chordate; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Musineae;	
1 (bases 1 to 731)	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.	
Ms.	Gessel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.	The WashU-HHMI Mouse EST Project Unpublished (1996)
Contact:	Marra M/Mouse EST Project	
WashU-HHMI Mouse EST Project		
Washington University School of Medicine		
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
Tel.: 314 286 1800		
Fax: 314 286 1810		
Email: mouseest@wustl.edu		
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.		
MG:376551		
Seq primer: -28M13 rev1 from Amersham		
High quality sequence stop: 428.		
Location/Qualifiers		
1..731		
/organism="Mus musculus"		
/strain="C57BL/6J"		
/note="Vector: PCMV-SPORT2; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. 13.5dpc embryos. PCMV-SPORT2 vector."		
/db_xref="taxon:10090"		
/clone="615727"		
/dev_stage="13.5dpc embryos"		
/lab_host="DH10B"		
<1..>731		
BASE COUNT	193 a 165 c 183 g 189 t 1 others	
ORIGIN		
Query Match	27.2%; Score 502; DB 23; Length 731;	
Best Local Similarity 89.9%; Pred. NO. 0.00e+00;		
Matches 570; Conservative 0; Mismatches 63; Indels 1; Gaps 1		
Db	95 CATGCCCTATATTGATATGATGTGAAGTAACCTTGACTTCAGAAGATGTCTGTGGAGCCCAA 154	
OY	335 CATTGCCTATATTGCAACAAGATGTGAACCTTGACTTCAGAAGATGTCTGTGGAGCCCAA 394	
Db	155 ACGCAGCACTTTAAGTGTGCAAGTAGAGTGAAGTGTGCAAGATCATTTTGTCCGAAA 214	
OY	395 ACGCAGTACCCTTAAGTGTGCAAGTAGAGTGAAGTGTGCAAGATCATTTTGTCCGAAA 454	
Db	215 CTCAAAGCAGATGTGACAGTGGCATCCCTGTGATTTGCTGCCAATATGATACCGTGGCAC 274	
OY	455 CTCAAAGCAGACATCTCTGGGGTTCCTCATTTGCTGCCAATATGATACGTGGCAC 514	
Db	275 TTTTGAATGGCTAGGGTCTCTGTGAAGTTCCTCTGTACCTGCCATCCATAAGCACTA 334	
OY	515 CTTGAGATGGCCAAGGTTCTCTGTGAATTTCTCTCTTCACTCTGTGCAATAGCACTA 574	
Db	335 CAGCATCATAGTGGCAAGATTGTGTCAGCAAGAACCTTACTGTCTGAGTGTACG 394	
OY	575 TAGCCTCGTTCAAGTGGCAAGAGTTGCTGGCCAGAACTCTACTGTCTTGAGCATCTGCG 634	
Db	395 TGCCAGCTCAGGCTCTGCTGACTTTGAGCAGCTGGAACAGATCCTGGAAGCTAT 454	
OY	635 TGCCAGCTCAGGCAAGGCTCTTCTGCACTTTGAGCAGCTGGAACAGATCCTGGAAGCTAT 694	
Db	455 TCCCCAATCAATATATATGCTCGGAGTGGCTTAAGGCTACTCTGAACATTTTGTGA 514	
OY	695 TCCCCAGGTGAAGTATATATGCTCGGAGTGGCAAAATGGCTACTCTGAACATTTTGTGA 754	

	Db	515	ATTTCCTAAAGATGTACGGAAAGCATTCCGCCAACACACCATCATGCGCAGAATAATGTGCT	574
	Oy	755	ATTTCCTAAAGATGTACCGAAGCGGCTTCCCCAGCACACATCATGCGCAGAATAATGTGCT	814
	Db	575	AACAGAGAGATGTAGAGAAAGTTAATCCTTTCTGGGGCTGCATCATCAACAAGTGGGAAA	634
	Oy	815	AACAGAGAGATGTAGAGAAAGTTAATCCTTTCTGGGGCTGCATCATCAACAAGTGGGAA	874
	Db	635	TGTCCTGCTGCTCTGTCTGTACAACTTCGAAGAAAACTGAGTTNGTATC-ACAGCTCAG	693
	Oy	875	TGGCGCAGGACTCTGTGTGTACTACTCGGAGAAAACTGAGTGGGGTGATCCACAGCTCAG	934
	Db	694	TGCACTGATGAGTGTTCAGATGCTGCTCAGGCC	727
	Oy	935	CGCACTGATGAGTGTTCAGATGCTGCTCAGGCC	968
RESULT	LOCUS	5	AA009421	494 bp mRNA EST 01-FEB-1997
DEFINITION			ze82c05.tl Soares fetal heart NBH19W Homo sapiens cDNA clone sequence.	
ACCESSION			AA009421	
NID			91470619	
KEYWORDS			EST.	
SOURCE			human.	
ORGANISM			Homo sapiens	
REFERENCE			Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrate; Eutheraia; Primates; Catarrhini; Homidae; Homo.	
AUTHORS			1 (bases 1 to 494) Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,D., Riklin,L., Ronlfing,T., Tan,F., Trevaskis,E., Waterson,R., Williamson,A., Wohldmann,P. and Wilson.R.	
TITLE	JOURNAL		Washington Merck EST Project Unpublished (1995)	
COMMENT			Contact: Wilson RK Washington Merck EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 638 Std Error: 0.00 Seq primer: mob.REGA+ET. Location/Qualifiers 1..494	
FEATURES	SOURCE		/organism="Homo sapiens" /note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGTAAGTGGAGCGCGCCGATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Felina Bonardo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBH19W."	
			/db_xref="taxon:9606" /clone_id="365480" /clone_lib="Soares fetal heart NBH19W" /sex="unknown" /dev_stage="19 weeks" /lab_host="DH10B (ampicillin resistant)" <1..>494	
BASE COUNT			140 a 126 c 117 g 111 t	
ORIGIN				

Query Match 26.6%; Score 491; DB 23; Length 494;
 Best Local Similarity 99.8%; Pred. No. 0.00e+00;
 Matches 492; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 2 GTGGAACATACCATCCGAGACATCTAGAGAGATCCGCTCTAGCTGTACTGTATGTGGGA 61
 |||||||
 QY 1236 GTGGAACATACCATCCGAGACATCTAGAGAGATCCGCTCTAGCTGTACTGTATGTGGGA 1235
 |||||||
 Db 62 GCAGCTAAGCTCAAAAGAGTGGAGCAGAGAACTACCTTCATCCGAGTACCAGCAGAGGTG 121
 |||||||
 QY 1296 GCAGCTAAGCTCAAAAGAGTGGAGCAGAGAACTACCTTCATCCGAGTACCAGCAGAGGTG 1355
 |||||||
 Db 122 AATCAATCTCACTGAGGAGGCTGCTAGACCTGAGAGAGTGTACCTTCCAGCAGCAGCT 181
 |||||||
 QY 1356 AATCAATCTCACTGAGGAGGCTGCTAGACCTGAGAGAGTGTACCTTCCAGCAGCAGCT 1415
 |||||||
 Db 182 ACTGTACCATGGGCGATCCCAAGTGGGGTCCCTACACCCATCCAGCTACTGAGCTGTCTA 241
 |||||||
 QY 1416 ACTGTACCATGGGCGATCCCAAGTGGGGTCCCTACACCCATCCAGCTACTGAGCTGTCTA 1475
 |||||||
 Db 242 TTACTTGTCTATTTCTCTGTCTCTACTCTGAGGGCTCTGAGTAATCTGTACTTCT 301
 |||||||
 QY 1476 TTACTTGTCTATTTCTCTGTCTCTACTCTGAGGGCTCTGAGTAATCTGTACTTCT 1535
 |||||||
 Db 302 CTATCTGACACACAAATGCCCAAGGCACTGCTGGGAGAGAAAGCAAGCAAGCAAG 361
 |||||||
 QY 1536 CTATCTGACACACAAATGCCCAAGGCACTGCTGGGAGAGAAAGCAAGCAAGCAAG 1595
 |||||||
 Db 362 TCTGAGAAATGATGACAGAAATCAATGGGAATCTGGGGACCCACACAAATCTCTGA 421
 |||||||
 QY 1596 TCTGAGAAATGATGACAGAAATCAATGGGAATCTGGGGACCCACACAAATCTCTGA 1655
 |||||||
 Db 422 AGATTATTTAAAGAAAAGATGCTGATTTGTATCATTAATCTTTACATGGCTTGTCTA 481
 |||||||
 QY 1656 AGATTATTTAAAGAAAAGATGCTGATTTGTATCATTAATCTTTACATGGCTTGTCTA 1715
 |||||||
 Db 482 GAGGAGGCAAGCT 494
 |||||||
 QY 1716 GAGGAGGCAAGCT 1728
 |||||||

RESULT 6
 LOCUS N28887 501 bp mRNA EST 04-JAN-1996
 DEFINITION yx63a11.r1 Homo sapiens cDNA clone 266396 5' similar to gb:M24470
 GMP REDUCTASE (HUMAN);.
 ACCESSION N28887
 NID g1147123
 KEYWORDS EST.
 SOURCE human clone=266396 primer=T7 library=Soares melanocyte 2NBHx
 vector=pt73D (Pharmacia) with a modified polylinker host=DH10B
 (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI Male. 1st strand
 cDNA was primed with a Not I - oligo(dT) primer
 [5'-TGTTTCAATCTGAGTGGAGCGGCGCGGCTTTTCTTTTCTTTT-3'],
 double-stranded cDNA was size selected, ligated to Eco RI adapters
 (Pharmacia), digested with Not I and cloned into the Not I and Eco
 RI sites of a modified pT73 vector (Pharmacia). Library
 constructed by Bento Soares and M.Fatima Bonafide. RNA from normal
 foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P.
 Albino.

ORGANISM Homo sapiens

Eukaryote; Metazoa; Eumetazoa; Bilateria; Coelomata;
 Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
 Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
 Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

AUTHORS

Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.

TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1995)

Contact: Wilson RK
 Washu-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence stops: 372
 Source: IMAGE Consortium, LNLN
 This clone is available royalty-free through LNLN; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 FEATURES
 location/Qualifiers
 1..501
 /organism="Homo sapiens"
 /clone="266396"
 <1..>501

BASE COUNT 120 a 116 c 130 g 134 t 1 others
 ORIGIN

Query Match 26.5%; Score 489; DB 19; Length 501;
 Best Local Similarity 99.4%; Pred. No. 0.00e+00;
 Matches 497; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 1 TCCTTTTCATTTTCGGAAGCTCAAAAGCAGACATCTCTGGGTTCCCATCATTTGCTGCCAAT 60
 |||||||
 QY 438 TCCTTTTCATTTTCGGAAGCTCAAAAGCAGACATCTCTGGGTTCCCATCATTTGCTGCCAAT 497
 |||||||
 Db 61 ATGATATCTGTGGGACCTTTGAGATGGCCAAAGGTTCTGTAGTTCCTCTCTCTACT 120
 |||||||
 QY 498 ATGATATCTGTGGGACCTTTGAGATGGCCAAAGGTTCTGTAGTTCCTCTCTCTACT 557
 |||||||
 Db 121 GATGTCATTAAGACATATAGCTCTGTTCACTGAGCAAGATTTGCTGGCCAGAAATCTGAT 180
 |||||||
 QY 558 GCTGTCCATTAAGACATATAGCTCTGTTCACTGAGCAAGATTTGCTGGCCAGAAATCTGAT 617
 |||||||
 Db 181 TGTCTTGAAGCATCTGGCTGCGACAGTCAAGGCAAGGCTTCTGACTTTGAGCAGCTGAA 240
 |||||||
 QY 618 TGTCTTGAAGCATCTGGCTGCGACAGTCAAGGCAAGGCTTCTGACTTTGAGCAGCTGAA 677
 |||||||
 Db 241 CAGATCCGGAAGCTATTTCCCGACAGTGAATATATATGCTGATGTTGGCAAAATGGCTAC 300
 |||||||
 QY 678 CAGATCCGGAAGCTATTTCCCGACAGTGAATATATATGCTGATGTTGGCAAAATGGCTAC 737
 |||||||
 Db 301 TCTGAACACTTTTGTGAATTTGTAAGATGTACGGAAGGCTTCCCGACACACCATC 360
 |||||||
 QY 738 TCTGAACACTTTTGTGAATTTGTAAGATGTACGGAAGGCTTCCCGACACACCATC 797
 |||||||
 Db 361 ATGCGAGGGAATGTGTACAGAGAGATGTAGAAAGCTCATCTTTCTGGGGCTGAC 420
 |||||||
 QY 798 ATGCGAGGGAATGTGTACAGAGAGATGTAGAAAGCTCATCTTTCTGGGGCTGAC 857
 |||||||
 Db 421 ATCATCAAGGGGAATTTGGGCGCAGGCTGTGTACTACTCGGAAGAAATCGAGAGTG 480
 |||||||
 QY 858 ATCATCAAGGGGAATTTGGGCGCAGGCTGTGTACTACTCGGAAGAAATCGAGAGTG 917
 |||||||
 Db 481 GGGTATCC-CAGCTCAGCGC 499
 |||||||
 QY 918 GGGTATCCACAGCTCAGCGC 937
 |||||||

RESULT 7
 LOCUS N30305 546 bp mRNA EST 05-JAN-1996
 DEFINITION yw69hd06.s1 Homo sapiens cDNA clone 257531 3'.
 ACCESSION N30305
 NID g1148825
 KEYWORDS EST.

SOURCE human clone=257531 primer=ml3 -40 forward 1library=Soares placenta
 8to9weeks 2bDH8to9w vector=PT73D (Pharmacia) with a modified
 polylinker host=DH10B (ampicillin resistant) Rsite1=Not I
 Rsite2=Eco RI two placentae: one from 8 weeks and another from 9
 weeks post conception. 1st strand cDNA was primed with a Not I -
 oligo(dT) primer
 [5'-TGTTTCAATCTGAGTGGAGCGGCGCGATTTTCTTTTCTTTT-3'],


```

CP 1595 CTGTTTCCTTCCTGCTCCCTCCAGAGAGTGGCTTTGGTGTGTCAGATAG 1536
DB 121 AGAAGTACAGAGTACTGACAGAGCCCTCAGAGTGTAGACACAGAAATGACAAATGAA 180
CP 1535 AGAAGTACAGAGTACTGACAGAGCCCTCAGAGTGTAGACACAGAAATGACAAATGAA 1476
DB 181 TACAGACCTCAGTACTGGAGTGGTGTAGAGACCCACTTGGATGCCCATGTGTAGAGT 240
CP 1475 TACAGACCTCAGTACTGGAGTGGTGTAGAGACCCACTTGGATGCCCATGTGTAGAGT 1416
DB 241 ACTGTCCTTGGAGGTGAGTACTGCTCAGTGTAGACACCCCTCAGTGTAGATGAT 300
CP 1415 ACTGTCCTTGGAGGTGAGTACTGCTCAGTGTAGACACCCCTCAGTGTAGATGAT 1356
DB 301 CACCTGCTGGTGAAGTGTAGAGTACTGCTCCTGCTCAACTCTTGTAGACTTGTGCTGC 360
CP 1355 CACCTGCTGGTGAAGTGTAGAGTACTGCTCCTGCTCAACTCTTGTAGACTTGTGCTGC 1236
DB 361 TCCCATATAGTACAGTACAGAGCGGATCCCTCAGATGTCTGGATGTATGTTCCAC 420
CP 1295 TCCCATATAGTACAGTACAGAGCGGATCCCTCAGATGTCTGGATGTATGTTCCAC 1236
DB 421 ATCTCCTTAAAGAGACTTCCACTGTCTTCCCTGTAGAGCTGTGATCTCA 472
CP 1235 ATCTCCTTAAAGAGACTTCCACTGTCTTCCCTGTAGAGCTGTGATCTCA 1184

RESULT 9
LOCUS N35234 474 bp mRNA EST 16-JAN-1996
DEFINITION XY21405.s1 Homo sapiens cDNA clone 271881 3'.
ACCESSION N35234
NID g1156376
KEYWORDS EST.
SOURCE human clone=271881 primer-m13 -40 forward library=Soares melanocyte
2bhm vector-p1773p (Pharmacia) with a modified poly1mer
host-DH10B (ampicillin resistant) Rsite1-Not I Rsite2-Eco RI Male.
1st strand cDNA was primed with a Not I - Oligo(dT) primer
[5'-TGTTCACATCTGAGTGGAGCGGCCGACAGTTTCTTTTCTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI adapters
(Pharmacia), digested with Not I and cloned into the Not I and Eco
RI sites of a modified p1773 vector (Pharmacia). Library
constructed by Bento Soares and M.Patima Bonaldo. RNA from normal
foreskin melanocytes (FS314) was kindly provided by Dr. Anthony P.
Albino.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonadata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 474)
REFERENCE Haller,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
AUTHORS Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The Washu-Merck EST Project
Unpublished (1995)
TITLE Contact: Wilson RK
JOURNAL Washu-Merck EST Project
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 384
Source: IMAGE Consortium, LNLN
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Location/Qualifiers
1. 474
FEATURES
Source /organism="Homo sapiens"

```

```

BASE COUNT 118 a 101 c 117 g 137 t 1 others
ORIGIN
Query Match 25.28; Score 465; DB 19; Length 474;
Best Local Similarity 99.48; Pred. No. 0.00e+00;
Matches 472; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

DB 1 TTGATATGTGAAGTCCATTTAGTGAACGGATTAAACAAATGATCTTAAAGCT 60
CP 1783 TTGATATGTGAAGTCCATTTAGTGAACGGATTAAACAAATGATCTTAAAGCT 1724
DB 61 GCCTCCTGAGACCAAGCCATGTAAAGATTATGACCAATCAGCATCTTCTTT 120
CP 1723 GCCTCCTGAGACCAAGCCATGTAAAGATTATGACCAATCAGCATCTTCTTT 1664
DB 121 AATATATCTTCAAGATGTGTGTGGTCCCAAGATTTCCATTGATTTCTTGATCAT 180
CP 1663 AATATATCTTCAAGATGTGTGTGGTCCCAAGATTTCCATTGATTTCTTGATCAT 1604
DB 181 TTTCAGACGTGTTGCTTCCCTGCTTCCCTCCAGATGAGTGCCTGGN-ATTGTTGTGT 239
CP 1603 TTTCAGACGTGTTGCTTCCCTGCTTCCCTCCAGATGAGTGCCTGGN-ATTGTTGTGT 1544
DB 240 GCAGATAGAGAGTACAGATTTACTGCAGAGCCCTCAGAGTGTAGACAGAAATGA 299
CP 1543 GCAGATAGAGAGTACAGATTTACTGCAGAGCCCTCAGAGTGTAGACAGAAATGA 1484
DB 300 CAAATATATACAGAGCTGTGAGTGTGGTGTAGAGACCCCATTTGGATGCCCAT 359
CP 1483 CAAATATATACAGAGCTGTGAGTGTGGTGTAGAGACCCCATTTGGATGCCCAT 1424
DB 360 GGTAGTACTGGTGGCTTGGAGGTAGAACTCTCAGATGACAGGCTCAGTAAG 419
CP 1423 GGTAGTACTGGTGGCTTGGAGGTAGAACTCTCAGATGACAGGCTCAGTAAG 1364
DB 420 ATTGATTCACCTCTGCTGGTACCTCGGATGAGTGTGTTCTCTGCTCACTCT 474
CP 1363 ATTGATTCACCTCTGCTGGTACCTCGGATGAGTGTGTTCTCTGCTCACTCT 1309

RESULT 10
LOCUS AA456771 478 bp mRNA EST 06-JUN-1997
DEFINITION zw27f02.r1 Soares ovary tumor NBH07 Homo sapiens cDNA clone 770523
ACCESSION AA456771
NID g2177192
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae;
Homo.
1 (bases 1 to 478)
REFERENCE Haller,L., Allen,M., Bowles,L., Dubuque,T., Gelsel,G., Jost,S.,
AUTHORS Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Schellenberg,K., Stepcoe,M., Tan,F., Theisling,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
The Washu-Merck EST Project 1997
Unpublished (1997)
TITLE Contact: Wilson RK
JOURNAL Washu-Merck EST Project
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amer sham.
Location/Qualifiers
FEATURES

```

1. 478
/organism="Homo sapiens"
/note="Organ: ovary; Vector: p1773D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - Oligo(dt) primer (5' TGTTCAACATCGAAGTGGAGCGGGCGGGTGTGTTTTTTTTTTTTTTT 3') / double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and Eco RI sites of the Not I and Eco RI sites of a modified p1773 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bernaldo."

/db_xref="taxon:9606"
/clone="770523"
/clone_1lb="Soares ovary tumor NBH0T"
/sex="Female"
/tissue_type="Ovarian tumor"
/lab_host="DH10B (ampicillin resistant)"
<1. >478
/db_xref="GDB:5980396"
/db_xref="NCBI:5980396"

BASE COUNT 131 a 122 c 122 g 103 t
ORIGIN

Query Match 25.2%; Score 465; DB 7; Length 478;
Best Local Similarity 99.8%; Pred. No. 0.00e+00;
Matches 466; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

12 TGGGGCGTGCTGAGTACAGAGCCTCAGAGGGAAGACAGTGAAGTCTTTTAAAG 71
1172 TGGGGCGTGCTGAGTACAGAGCCTCAGAGGGAAGACAGTGAAGTCTTTTAAAG 1231

Db 72 AGATGTGGAACATACCATCCGAGACATCCTAGAGAGGATCCGCTTACGTACTTATGT 131
1232 AGATGTGGAACATACCATCCGAGACATCCTAGAGAGGATCCGCTTACGTACTTATGT 1291

Db 132 GGGAGCAGCTAAGCTCAAGAGTGTAGCAGAGAGAACTACCTATCCGAGTACCCAGCA 191
1292 GGGAGCAGCTAAGCTCAAGAGTGTAGCAGAGAGAACTACCTATCCGAGTACCCAGCA 1351

Db 192 GGTGAATCCATCTTCAGTAGAGCGTGTAGACCTGAGCAAGTCTTACCTCCCAAGGCAC 251
1352 GGTGAATCCATCTTCAGTAGAGCGTGTAGACCTGAGCAAGTCTTACCTCCCAAGGCAC 1411

Db 252 CAGTACTTACCATTTGGGGCATCCCAAGTGGGTCCTACCCATCCAGTACTGACGCTC 311
1412 CAGTACTTACCATTTGGGGCATCCCAAGTGGGTCCTACCCATCCAGTACTGACGCTC 1471

Db 312 TGTATTACTTTGTCATTTCTGTCGTCACCTGAGGGGCTCCTGCAGTACTCTGAC 371
1472 TGTATTACTTTGTCATTTCTGTCGTCACCTGAGGGGCTCCTGCAGTACTCTGAC 1551

Db 372 TTCTCTATTTGCACACAAATATGCCCAAGGACCTCCTGGGAGGAAGAAGAACAA 431
1532 TTCTCTATTTGCACACAAATATGCCCAAGGACCTCCTGGGAGGAAGAAGAACAA 1591

Db 432 ACAGTGTGAGAAAATGATGCAGAAAATCAATATGGGATCTGGGGAC 478
1592 ACAGTGTGAGAAAATGATGCAGAAAATCAATATGGGATCTGGGGAC 1638

RESULT 11
LOCUS AA524317 599 bp mRNA EST 21-AUG-1997
DEFINITION ng33h05.s1 NCI-CGAP_C03 Homo sapiens cDNA clone IMAGE:936633, MRNA
ACCESSION AA524317
NID g255245
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryota; Eukaryota; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;
Homo.
1 (bases 1 to 599)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TYPE	Journal Comment
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	unpublished (1997)
Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arraying: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/bdrp/image/Imags.html	
FEATURES	
source	<p>Insert length: 590 Std Error: 0.00 Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop: 275. Location/Qualifiers</p> <p>1..599</p> <p>/organism="Homo sapiens" /note="Vector: pUT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - Oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pUT3 vector. Library went through one round of normalization." /db_xref="taxon:9606" /clone_image="IMAGE:936633" /clone_lib="NCI CGAP_C03" /sex="pooled" /tissue_type="colon" /lab_host="DH10B" <1..>599</p>
BASE COUNT	156 a 129 c 142 g 171 t 1 others
ORIGIN	
Query Match	25.0%; Score 460; DB 9; Length 599;
Best Local Similarity	97.7%; Pred. No. 0.00e+00;
Matches	501; Conservative 0; Mismatches 6; Indels 6; Gaps 6;
1	ACAGCCAATGAGCTCTCTCTTATGAGGTAATTAATATACAAACACTCAGAGCTTT 60
1843	ACAGCCAATGAGCTCTCTCTTATGAGGTAATTAATATACAAACACTCAGAGCTTT 1784
61	TTAGATATGTGAAGTCCCAATTTAGTGAAGCGGATTAACAACATGATCTAAAGCCT 120
1783	TTAGATATGTGAAGTCCCAATTTAGTGAAGCGGATTAACAACATGATCTAAAGCCT 1724
121	GCCTCCTCAGACCAAGGCATGTAAAGATTATGTACCAATCAGCATCTTTCTTTT 180
1723	GCCTCCTCAGACCAAGGCATGTAAAGATTATGTACCAATCAGCATCTTTCTTTT 1664
181	AATTAATCTCAGAGATGTGTGTGGTCCCAAGATATCCATTGATTTCTTGATCATTT 240
1663	AATTAATCTCAGAGATGTGTGTGGTCCCAAGATATCCATTGATTTCTTGATCATTT 1604
241	TTCCGAGACTGTTTCTTCTTCTTCCCTCCCACTGAGTGCCTTGGCATTTTGTGTGT 300
1603	TTCCGAGACTGTTTCTTCTTCTTCCCTCCCACTGAGTGCCTTGGCATTTTGTGTGT 1544
301	GCAGATAGGAAGTACAGAGTTACTGCAAGAGCCCTCAGAGTGAACACAGAAATGA 360
1543	GCAGATAGGAAGTACAGAGTTACTGCAAGAGCCCTCAGAGTGAACACAGAAATGA 1484
361	NCAGAGTAATACAGACTGCAGTACGTGGGATGGGAGAGACCCCACTGGATGCCCAT 420
1483	NCAGAGTAATACAGACTGCAGTACGTGGGATGGGAGAGACCCCACTGGATGCCCAT 1424
421	GGTAGAGTACTGTGC-TTGGAGGGTA-AACTGCTCAGGTC-AGCACGCC-CAGTGAG 476

Matches	459;	Conservative	0;	Mismatches	4;	Indels	1;	Gaps	1;
Db	1	ACAGCAATGAGCTCTCTCTATTATGAGGATTTCAATATATCAAAACACTTGAGGCTTT	60						
Cp	1843	ACAGCAATGAGCTCTCTCTATTATGAGGATTTCAATATATCAAAACACTTGAGGCTTT	1784						
Db	61	TTAGATATGTGAAGGTCCCAATTAGTGAAGCGGATTAACAAACATGATTCGAAAGGCT	120						
Cp	1783	TTAGATATGTGAAGGTCCCAATTAGTGAAGCGGATTAACAAACATGATTCGAAAGGCT	1724						
Db	121	GCCTCCCTAGACCAAGGCCATGTAAAGATTATATGACCAATCAGCATCTTTCCCTTT	180						
Cp	1723	GCCTCCCTAGACCAAGGCCATGTAAAGATTATATGACCAATCAGCATCTTTCCCTTT	1664						
Db	181	AATAATCTTCAGGATGTTGTGTGGGTCCCAAGATTCACATTTGATTTTCTGCATCAT	240						
Cp	1663	AATAATCTTCAGGATGTTGTGTGGGTCCCAAGATTCACATTTGATTTTCTGCATCAT	1604						
Db	241	TTCTCAGACTGTGTCCTTCCTTCCTCCCTCCCAAGTGGCTTGAGATTTTGTGCT	300						
Cp	1603	TTCTCAGACTGTGTCCTTCCTTCCTCCCTCCCAAGTGGCTTGAGATTTTGTGCT	1544						
Db	301	GCAGATAGAGAGTACAGAGTTACTACTGCAGAGACCCTCAGAGATGAGACAACAGAAATGA	360						
Cp	1543	GCAGATAGAGAGTACAGAGTTACTACTGCAGAGACCCTCAGAGATGAGACAACAGAAATGA	1484						
Db	361	CAAACTAATACAGAGCTGCAGTACTGCTGGGAGTGGTGAAGACC-ACCTGGATGCCCAT	419						
Cp	1483	CAAACTAATACAGAGCTGCAGTACTGCTGGGAGTGGTGAAGACC-ACCTGGATGCCCAT	1424						
Db	420	GGTAGAGTACTGGTGCCTTGGGAGGTGGAATGCTCAGCTTA	463						
Cp	1423	GGTAGAGTACTGGTGCCTTGGGAGGTGGAATGCTCAGCTTA	1380						
RESULT	14								
LOCUS	AA616194	642 bp	mRNA	EST	07-OCT-1997				
DEFINITION	vo96g93.1	Soares mouse mammary gland NbMg Mus musculus cDNA clone							
	1066996	5 similar to gb:M24470 GMP REDUCTASE (HUMAN); mRNA							
ACCESSION	AA616194								
NID	92503399								
KEYWORDS	EST.								
SOURCE	house mouse.								
ORGANISM	Mus musculus								
REFERENCE	Eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;								
AUTHORS	Rodentia; Scurionathi; Muridae; Murinae; Mus.								
	1 (bases 1 to 642)								
	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,								
	Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,								
	Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,								
	Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and								
	Waterson, R.								
TITLE	The Mashu-HHMI Mouse EST Project								
JOURNAL	Unpublished (1996)								
COMMENT									
	Contact: Marra M/Mouse EST Project								
	Mashu-HHMI Mouse EST Project								
	Washington University School of Medicine								
	444 Forest Park Parkway, Box 8501, St. Louis, MO 63108								
	Tel: 314 286 1800								
	Fax: 314 286 1810								
	Email: mouseest@wustl.edu								
	This clone is available royalty-free through LLNL; contact the								
	IMAGE Consortium (info@image.llnl.gov) for further information.								
	MG:589356								
	Seq primer: -28m13 rev2 ET from Amersham								
	High quality sequence stop: 497.								
	Location/Qualifiers								
	1..642								

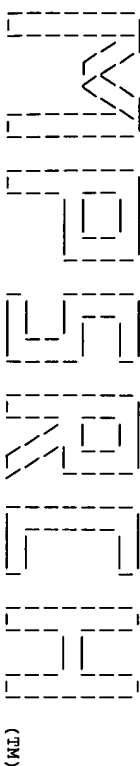
Query Match	Best Local Similarity	Score	DB	Length
Matches 501; Conservative	90.4%;	24.3%;	DB 10;	642;
			Pred. No. 0.00e+00;	
			Mismatches 53; Indels 0; Gaps 0;	
Db	89	CATGCTCATATTGTAATATGATGTGAAGAACTTGACTTCAAGATGTCCTGTTGAGGCCAA	148	
Qy	335	CATGCTCATATTGTCACACGATGTGAAGCTGAGCTTCACGAGATGTCCTTTTGAGGCCAA	394	
Db	149	ACGGACACCTCTTAAGTGTGAGGAGGTGAGCTACACAGATCATTTGCTCCGGAA	208	
Qy	335	ACGGATACCTTTAAGTGTGAGGAGGTGAGCTACACAGATCATTTGCTCCGGAA	454	
Db	209	CTCAAGACAGATGTACACATGCGATCCCTGCTATGCTGCTCCAAATGTGATACCGTGGCAC	268	
Qy	455	CTCAAGACAGATGTACATCTGGGGTTCCCATCATTTGCTGCGCAATGTGATGTGGGGCAC	514	
Db	269	TTTGAATGGCTAGGGTCTCTGTAACTTCCCTGTTACCTGCATCATTAAGCACTA	328	
Qy	515	CTTTGAGATGGCCAAAGGTTCTCTGTAACTTCTCTCTTCACTGCTGCTCAATGACACTA	574	
Db	329	CAGCATCATGATAGGCGAAGATTGTGACGACAAATCCGACGTCTGATGCTTACG	388	
Qy	575	TAGCTCTGTTAGTGGCAGAGGTTGCTGGCGCAAAATCCTGACTGTTGAGCATCTGGC	634	
Db	389	TGCCAGCTCAGAGCTCAGGCTCTGCTGACTTTGAGCAGCTGGAGACAGATCCGGAAGCTAT	448	
Qy	635	TGCCAGCTCAGGACAGGCTCTTCTGACTTTGAGCAGCTGGAGACAGATCCGGAAGCTAT	694	
Db	449	TCGCCAAGTCAATATATATATCCCTGATGTGGCTACGCGCTACTCTGAACTTTTGTGA	508	
Qy	695	TCGCCAGGTGAATATATATCCTGATGTGGCAAAATGGCTACTGTAACACTTTGTGTA	754	
Db	509	ATTGTAAAGATGATGAGGAGAGGATTCGCCACACACCAATCATGTGCAAGAAATGTGCT	568	
Qy	755	ATTGTAAAGATGATGAGGAGAGGCTTCGCCACACACCAATCATGTGCAAGAAATGTGCT	814	
Db	569	AACAGGAGAGATGTAGAGAGTAAATCCCTTTCTGGGGCTGACATCATCAAAAGTGGGAT	628	
Qy	815	AACAGGAGAGATGTAGAGAGTCAATCCCTTTCTGGGGCTGACATCATCAAAAGTGGGAT	874	
Db	629	TGGTCTCTGGCTCTG 642		
Qy	875	TGGGCAAGCTCTG 888		
RESULT	15			
LOCUS	AA830800	474 bp	mRNA	EST
DEFINITION	oc65d03.g1 NC1_CGAB_GCB1 Homo sapiens cdna clone IMAGE:1353701 3',			
ACCESSION	AA830800			
MID	g2303899			
KEYWORDS	EST.			
SOURCE	human.			

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 474)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.11nl.gov/dbirp/image/image.html
Insert Length: 1720 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham.
Location/Qualifiers
1. 474
/organism="Homo sapiens"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCGAAGTGGGAGCGGCCGCTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1353701"
/clone_lib="NCI-CGAP-GCBI"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
BASE COUNT 125 a 93 c 109 g 147 t
ORIGIN
Query Match 24.18; Score 445; DB 16; Length 474;
Best Local Similarity 98.7%; Pred. No. 0.00e+00;
Matches 461; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
Db 10 ACAGTCAATAGAGCTCTCTCTTATGTAGGTATTTCAATATACAAACACTTCAGAGCTTT 69
Cp 1843 ACAGTCAATAGAGCTCTCTCTTATGTAGGTATTTCAATATACAAACACTTCAGAGCTTT 1784
Db 70 TTAGATATGTGAAGTCCATTTAGTGAAGCGATTAAACAACATGATCTAAAGCCT 129
Cp 1783 TTAGATATGTGAAGTCCATTTAGTGAAGCGATTAAACAACATGATCTAAAGCCT 1724
Db 130 GCCTCCTTAGACCAAGCGCATTAAGATTTATGTACCAATCAGCATCTTTCCTTTT 189
Cp 1723 GCCTCCTTAGACCAAGCGCATTAAGATTTATGTACCAATCAGCATCTTTCCTTTT 1664
Db 190 AATAATCTTAGAGATGTGTGGTGGTCCCAATTTCCCATTTGATTTCTTGATCATTT 249
Cp 1663 AATAATCTTAGAGATGTGTGGTGGTCCCAATTTCCCATTTGATTTCTTGATCATTT 1604
Db 250 TTCTCAGACTGTTGCTTCTTCTCTCCAGTAGTGCCTGGGATATTGTGTGT 309
Cp 1603 TTCTCAGACTGTTGCTTCTTCTCTCCAGTAGTGCCTGGGATATTGTGTGT 1544
Db 310 GCAGATAGAGAGTACTGAGAGAGCCCTCAGAGTGTAGACAAACAGAAATGA 369

|||||
Cp 1543 GCAGATAGAGAGTACTGAGAGAGCCCTCAGAGTGTAGACAAACAGAAATGA 1484
Db 370 CAAGTAAATACAGAGCTGCAGTAGCTGGGATGGGTAGAGACCC-ACCTGGGATGCC-AT 427
Cp 1483 CAAGTAAATACAGAGCTGCAGTAGCTGGGATGGGTAGAGACCCACCTGGGATGCCAT 1424
Db 428 GGTAGAGTACTGGCTGGCTGGGAGGTAGAACTGCTCAGGTCTAGCA 474
Cp 1423 GGTAGAGTACTGGCTGGCTGGGAGGTAGAACTGCTCAGGTCTAGCA 1377

Search completed: Wed Jun 17 13:00:01 1998
Job time : 3135 secs.

This Page Blank (uspto)



(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Jun 15 15:56:34 1998; Maspar time 10.68 Seconds

Tabular output not generated. 525.802 Million cell updates/sec

Title: >US-08-774-169-1

Description: (1-366) from US08774169.pep

Sequence: 2632 1 MTCCLPALRFIATPRLSAMP.....RRTFIRVYQGVNPIFSEAC 366

Scoring table: PAM 150
Gap 11

Searched: 124785 seqs, 15338987 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseg31-2

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27

Statistics: Mean 34.191; Variance 147.545; scale 0.232

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	349	13.3	514	1	Human IMPDH.	3.84e-21
2	347	13.2	514	1	Chinese hamster IMPDH	5.87e-21
3	112	4.3	438	25	Staphylococcus aureus	1.28e+00
4	111	4.2	487	23	Sugar biosynthesis en	1.52e+00
5	101	3.8	168	27	E. coli colonisation	8.30e+00
6	97	3.7	661	9	Enzyme involved in ei	1.61e+01
7	97	3.7	661	20	Biosynthetic enzyme o	4.26e+01
8	91	3.5	256	19	Helicobacter pylori 2	3.63e+01
9	92	3.5	265	2	Mouse serine protease	4.26e+01
10	91	3.5	267	2	Sequence of serine pr	4.26e+01
11	92	3.5	422	7	ZmRL1 homologue prote	3.63e+01
12	92	3.5	520	25	Arabidopsis thaliana	5.00e+01
13	90	3.4	216	26	C. glutamicum orf3 pr	5.86e+01
14	89	3.4	307	26	Novel human phosphoy	5.86e+01
15	89	3.4	359	20	Prostaglandin DP rece	5.86e+01
16	89	3.4	558	7	XRS	5.86e+01
17	89	3.4	2408	2	Translation of ORF 3	5.86e+01
18	88	3.3	202	2	Glycopeptide resistin	6.87e+01
19	87	3.3	289	27	Rat syntaxin 3 protei	8.04e+01

20	87	3.3	343	5	R28887	pR18 fusion protein.	8.04e+01
21	88	3.3	403	13	R71933	Cobra venom factor C-	6.87e+01
22	87	3.3	530	5	R28888	R1 paraneoplastic ant	8.04e+01
23	86	3.3	750	1	P82940	Artrial natriuretic po	9.40e+01
24	85	3.2	211	18	W00581	Programmed mammalian	1.10e+02
25	85	3.2	211	12	R60867	RP-2 programmed cell	1.10e+02
26	84	3.2	255	6	R30808	3-alpha-hydroxy-ster	1.28e+02
27	83	3.2	286	17	R88917	Modified incp protein	1.50e+02
28	84	3.2	305	25	W23077	Sulfolobus solfataric	1.28e+02
29	85	3.2	376	15	R80016	Marmoset alpha-1,3-ga	1.10e+02
30	85	3.2	369	19	W04215	Rat vas deferens p2x	1.10e+02
31	85	3.2	399	14	W04218	Human urinary bladder	1.10e+02
32	83	3.2	421	14	R71416	C. glutamicum protein	1.50e+02
33	83	3.2	421	12	R63566	Feedback inhibition r	1.50e+02
34	83	3.2	421	23	W06586	B. lactofermentum asp	1.50e+02
35	83	3.2	421	15	R86307	Corynebacterium gluta	1.50e+02
36	84	3.2	491	18	R82243	Human fibrinogen B-be	1.28e+02
37	85	3.2	562	12	R63588	Full length HA protei	1.10e+02
38	84	3.2	572	8	R40843	Bilirubin oxidase.	1.28e+02
39	85	3.2	748	10	R52706	Human enkephalinase.	1.10e+02
40	85	3.2	750	10	R52707	Rat enkephalinase.	1.10e+02
41	85	3.2	750	10	R52707	Enkephalinase (rat).	1.10e+02
42	83	3.2	842	12	R63570	Feedback inhibition r	1.50e+02
43	83	3.2	842	12	R63567	Feedback inhibition r	1.50e+02
44	84	3.2	950	2	R06376	Product of the sscI g	1.28e+02
45	85	3.2	1266	24	W25157	Tomato immunity 2 (fu	1.10e+02

ALIGNMENTS

RESULT ID	1	ALIGNMENTS
AC	R05432; standard; protein; 514 AA.	
AC	R05432; 31-AUG-1990 (first entry)	
DE	Human IMPDH.	
KW	Inosine 5'-mono-phosphate dehydrogenase; hepatomas;	
OS	guanosine 5'-monophosphate.	
PN	MO9001545-A.	
PD	02-FEB-1990.	
PF	02-AUG-1989; U0344.	
PR	12-AUG-1988; US-232302.	
PA	(ARCH-) Arch. Dev. Corp.	
PI	Collart FR, Huberman E;	
DR	WPI: 90-083504/11.	
DR	P-PSDB; R05432.	
PT	DNA encoding eukaryotic inosine 5'-mono-phosphate dehydrogenase -	
PT	used to detect high levels of enzyme activity of tumour cells.	
PS	Claim 8; Fig 1; S1PP; English.	
CC	A 35 aa segment corresponds to deduced aa residues 336-370 in	
CC	both the human and Chinese hamster proteins.	
CC	See also Q03541.	
SQ	Sequence 514 AA:	

Query Match 13.3%; Score 349; DB 1; Length 514;

Best Local Similarity 30.3%; Pred. No. 3.84e-21;

Matches 71; Conservative 66; Mismatches 93; Indels 4; Gaps 4;

Db	231	rdylpaaekdkkqllcgaaigtbeddkyridllagagvddvvlidsqgnslfimmklyl 290	
Qy	104	OEFGAGPDCLEHLMASSGSSGDFEQLEILAIPOVKYICDVANGVSEHVEFVKDV 163	
Db	291	kdyrlpqlvlgvgnvtaagknllidagvdlrlygmssgscitqevlaagrpqatavkyv 350	
Qy	164	KRRPQHTIMAGNVYGEWEELILSGADIIKVGIGSVCTRRKRTGVGPOLSAVMEC 223	
Db	351	seyartrfyvviadggqigvghakalagastvmmsllaateapgevfdsqgtrlkx 410	
Qy	224	ADAHAGIKGHILISDGGSCGCDVAKAFAGADRVMLAGMLAGSSESGELIENDGKKYL 283	
Db	411	yrngmsldamdkhlsqnyr-fseadkikva-gg-vsgavqdk-gslkhfvpyl 460	

PS Claim 2; Fig 2; 4bp; English.

CC This is an isolated *Staphylococcus aureus* Coenzyme A disulphide reductase

CC (CoADR). CoADR catalyses the specific reduction of Coenzyme A disulphide

CC to Coenzyme A with the concomitant oxidation of NADP to NADP⁺. The DNA

CC encoding CoADR can be obtained from an *S. aureus* genomic DNA library

CC using PCR primers with the standard techniques. CoADR was fractionated

CC from cellular extracts of *S. aureus* by following the NADPH- and CoA

CC disulphide dependent reduction of dithio-bis-2-nitrobenzoic acid (DTNB).

CC The *S. aureus* CoADR can be used to identify compounds which can inhibit

CC the enzyme. Such compounds can be used to treat Gram-positive bacterial

CC infections such as staphylococcal and enterococcal infections. They can

CC be used for treating gastroenteritis, toxic shock syndrome, scalded skin

CC syndrome, dental infections, bone and joint infections, pneumonia and

CC emphysema, meningitis, cerebritis, endocarditis, bacteraemia, septic

CC shock, septicaemia, food poisoning, or enteritis. The products can also

CC be used for detection and diagnosis.

SO Sequence 438 AA;

CC

Query Match 4.3%; Score 112; DB 25; Length 438;

Best Local Similarity 27.9%; Pred. No. 1.28e+00;

Matches 34; Conservative 26; Mismatches 51; Indels 11; Gaps 8;

Db 145 angdxkylvvgagysvse-venlenerglhpcllnrsdtklnklmdamgpldelatre 203

Qy 243 PGDAKAFAGAGAFVMLGMLAGHSBG--GELIEDGKKRYLFGMSSEAMKRYAGCV 300

Db 204 lpyrlnealnagneltfksgkvevdydmlliegvgbphskflessnjk-ldrxg-flpy 261

Qy 301 AEYASSEG-KTV---EYPER-GVEHTTIDILG-GIRSCITVGAAKKEISRRTTFIRV 354

Db 262 nd 263

Qy 355 TQ 356

RESULT 4

ID W19740 standard; Protein; 487 AA.

AC W19740;

DT 18-SEP-1997 (first entry)

DE Sugar biosynthesis enzyme EryBI.

KW Polypeptide; glycosylation; eryBI; L-mycarose; antimicrobial;

KM antibiotic; antifungal; fungicide; anticancer; cytostatic;

KW anthelmintic.

OS Saccharopolyspora erythraea.

PN M09723630-A2.

PD 03-JUL-1997.

PF 23-DEC-1996; U20238.

PR 21-DEC-1995; US-576626.

PA (ABBO) ABBOTT LAB.

PI Donadio S, Katz L, Staver MJ, Summers RG;

DR WPI: 97-351066/32.

DR N-PSDB: T72685.

PT New genes involved in sugar biosynthesis and attachment - used to

PT generate polyclonal antimicrobials etc. with altered pattern of

PT glycosylation

PS Disclosure; Fig 4B; 85pp; English.

CC EryBI (W19740) is an enzyme involved in the biosynthesis of the

CC sugar L-mycarose. It is one of 10 enzymes (see also W19734-39 and

CC W19741-43) predicted to be involved in D-desosamine or L-mycarose

CC biosynthesis and attachment that have been identified from gene

CC clusters (T72684 and T72685) of *Saccharopolyspora erythraea*. Novel

CC glycosylation-modified polypeptides are produced by selectively

CC altering, inactivating or augmenting the eryB and/or eryC genes

CC encoding these sugar biosynthesis enzymes and introducing them into

CC polyclide-producing microorganisms.

SO Sequence 487 AA;

Query Match 4.2%; Score 111; DB 23; Length 487;

Best Local Similarity 28.4%; Pred. No. 1.52e+00;

Matches 21; Conservative 22; Mismatches 28; Indels 3; Gaps 3;

Db 203 pnfwtlvaqlrmlhndvnmlditvlaacvptavdrdradvarlpgsfgarllhs 262

CC Protective antigens (W05194-98) were identified by screening a DNA
CC library of H. pylori with serum from mice immunised with H. felis
CC sonicate and isolating E. coli clones expressing H. pylori proteins
CC recognised by anti-H. felis antibodies. Cloned proteins were
CC grouped into families (see also W05194-98), and representative cDNA
CC clones of Families C, E, G, H and B (T43043-47) were sequenced.
CC Recombinant protective antigens can be produced for use in vaccines.
SQ Sequence 256 AA;

Query Match 3.5%; Score 91; DB 19; Length 256;
Best Local Similarity 27.3%; Pred. No. 4.26e+01;
Matches 15; Conservative 14; Mismatches 23; Indels 3; Gaps 2;

Db 168 gnveaylllddyysgndqg-lepdkdxlyvykmaadmsrratyeqlaesyr 220
QY 251 GAGADFWLGGMLAGHSESGELIERDCKYKLFYGSSEMAKMK-YAGVAEYR 304

RESULT 9
ID R08269 standard; protein; 265 AA.
AC R08269;
DT 07-MAR-1991 (first entry)
DE Mouse serine protease precursor, deduced from the DNA.
KM Leukemia elastase; pulmonary emphysema; rheumatoid arthritis;
OS Mus musculus.
FH Key Location/Qualifiers
FT Protein 29..265
FN J0268680-A.
PD 02-NOV-1990.
PE 07-APR-1989; 088798.
PR 07-APR-1989; JP-088798.
PA (TORA) TOKAY IND INC.
DR WPI; 90-372006/50.
DR N-PSDB; Q06870.
PT Mouse model system - comprises mouse serine protease which shows
PT biological activity by polypeptide.
PS Claim 2; Page 1; 6pp; Japanese.
CC A mouse model system involving the serine protease can be used to
CC investigate the role of leukemia elastase in the development or
CC exacerbation of diseases such as pulmonary emphysema, rheumatoid
CC arthritis etc.
SQ Sequence 265 AA;

Query Match 3.5%; Score 92; DB 2; Length 265;
Best Local Similarity 23.2%; Pred. No. 3.63e+01;
Matches 19; Conservative 30; Mismatches 26; Indels 7; Gaps 6;

Db 185 nvccllyprirgagifdgsgpvlvcnmlvqglds-flr-ggcg-sgllpdatapvaeftdw 241
QY 202 SVCT--TRKKTGVGYPQLSAVMECADAHGKHIIISDGCSCPDVAKAGAGADPVM- 258

Db 242 Insllrshndhllhpkdreg 263
QY 259 LGGMLAGHSESG-GLIERDCK 279

RESULT 10
ID P80335 standard; protein; 267 AA.
AC P80335;
DT 04-OCT-1990 (first entry)
DE Sequence of serine protease (SP) of human myeloid cell
DE origin and leader peptide
KM Serine protease; myeloid cell; intravascular coagulation treatment;
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..29
FN P80806621-A.
PD 07-SEP-1988.
PR 26-FEB-1988; J00205.

PR 09-SEP-1987; JP-225540.
PA (TORA) Toray Ind Inc (AOXI/).
PI Aoki Y, Okano K, Naruto M, Shimizu H, Nakamura H;
DR WPI; 88-271161/38.
DR N-PSDB; N80315.
PT Serine protease of human myeloid cell origin -
PT useful in disseminated intravascular coagulation treatment
PS Disclosure; Fig 6; 52pp; Japanese.
CC A SP of human myeloid cell origin is new. Also new are precursors of SP
CC having an N-terminal splittable or signal peptide; DNA sequences coding
CC for SP and precursors and transcription-controlling DNA sequence required
CC for expression of the gene. SP has antithrombotic activity, useful in
CC disseminated intravascular coagulation. By expression of protease gene
CC in suitable transformant organism SP can be obtained in large quantity.
CC SP DNA selected and cloned from a library of human myeloid cell origin,
CC inserted in suitable vector and expressed in transformant by a suitable
CC organism such as E.coli HMS-174 or HB-101 a yeast.
SQ Sequence 267 AA;

Query Match 3.5%; Score 91; DB 2; Length 267;
Best Local Similarity 30.5%; Pred. No. 4.26e+01;
Matches 18; Conservative 19; Mismatches 17; Indels 5; Gaps 5;

Db 184 snvccllyprgagifdgsgpvlvcnmlvqglds-flr-ggcg-sgllpdatapvaeftdw 239
QY 201 GSVC-TR-KKTGVGYPQLSAVMECADAHGKHIIISDGCSCPDVAKAGAGADPV 257

RESULT 11
ID R39819 standard; protein; 422 AA.
AC R39819;
DT 12-JAN-1994 (first entry)
DE ZmPK1 homologue protein kinase.
KM Kinase; homologue; differentiation; prolongation; proliferation;
KW ss.
OS Solanaceae.
PN J05161494-A.
PD 28-JUN-1993.
PE 13-DEC-1991; 330418.
PR 13-DEC-1991; JP-330418.
PA (SUMO) SUMITOMO CHEM CO LTD.
DR WPI; 93-239258/30.
DR N-PSDB; Q47060.
PT ZmPK1 homologue protein kinase gene from solanaceae - used for
PT control of differentiation, prolongation and proliferation of
PT plant
PS Claim 1; Page 5-6; 6pp; Japanese.
CC ZmPK1 homologue protein kinase or its gene is useful for the control
CC of differentiation, prolongation and proliferation of plants. The
CC gene can also be used to study agents which enhance or inhibit ZmPK1
CC homologue protein kinase.
SQ Sequence 422 AA;

Query Match 3.5%; Score 92; DB 7; Length 422;
Best Local Similarity 29.8%; Pred. No. 3.63e+01;
Matches 17; Conservative 19; Mismatches 19; Indels 2; Gaps 2;

Db 304 felitahpqnimemyvnlamsdgydelldkklvgtslegv-tsiaaiah-kci 358
QY 88 FSLFTAVHKHYSLVQWQEFAGQNPDCLEHLAASSGSSPFDQLEQLLEHLPQVKYI 144

RESULT 12
ID W26640 standard; protein; 520 AA.
AC W26640;
DT 29-JAN-1998 (first entry)
DE Arabidopsis thaliana ferulate-5-hydroxylase.
DE ferulate-5-hydroxylase; lignin; delignification; pulp; paper;
KM feedstock; transgenic plant; alfalfa; Medicago; rice; Oryza;
KM maize; Zea mays; oilseed rape; Brassica; forage grass;
KM tobacco; Nicotiana; eucalyptus; pine; Pinus; spruce; Picea;
KM poplar; Populus; crop improvement; Arabidopsis thaliana.
OS Arabidopsis thaliana.

PN WO9723599-A2.
 PD 03-JUL-1997.
 PF 19-DEC-1996; U20094.
 PR 14-MAR-1996; US-013388.
 PR 22-DEC-1995; US-009119.
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 PA (PURD) PURDUE RES FOUND.
 PI Chapple C;
 DR WPI; 97-351042/32.
 DR N-PSDB; T90504-05.
 PT Nucleic acid encoding plant ferulate-5-hydroxylase - useful for
 PT altering guaiacyl:syringyl lignin monomer ratio in plants, to
 PT increase chemical delignification in paper and pulp making or to
 PS increase digestibility of feedstocks
 PS Claim 1; Page 27-28; 43pp; English.
 CC This polypeptide comprises Arabidopsis thaliana ferulate-5-
 CC hydroxylase (F5H), a key enzyme in lignin biosynthesis that
 CC catalyses the conversion of ferulate to 5-hydroxyferulate. Its
 CC amino acid sequence can be deduced from isolated cDNA and genomic
 CC DNA clones (see T90504-05). A claimed chimeric gene that causes an
 CC altered guaiacyl:syringyl lignin monomer ratio in a transformed
 CC plant comprises the F5H cDNA or genomic clone linked either in the
 CC sense or antisense direction to at least one regulatory sequence.
 CC The gene allows the alteration of a plant's lignin content.
 CC especially to increase the syringyl lignin content. This renders
 CC the lignin more susceptible to chemical delignification of use to
 CC the paper and pulp industries. Alteration of the lignin content of
 CC grasses increases the digestibility for animal feedstocks.
 CC Suitable host plants include alfalfa, rice, maize, oilseed rape,
 CC forage grasses, tobacco, and tree crops such as eucalyptus, pine,
 CC spruce and poplar.
 CC Sequence 520 AA;

Query Match 3.5%; Score 92; DB 25; Length 520;
 Best Local Similarity 26.5%; Pred. No. 3.63e+01;
 Matches 9; Conservative 16; Mismatches 7; Indels 2; Gaps 2;
 Db 249 fiddidehmkkengnavdgdvdtmdwd11 282
 156 FVMD-VKRPHTIMAGNVGEMVELI 187

RESULT 13
 ID W37716 standard; Protein; 216 AA.
 AC W37716;
 DT 12-MAR-1998 (first entry)
 DE C. glutamicum orf3 protein.
 KM lysG: lysE, ORF3; lysine transport; regulatory protein; export protein;
 KM Microbial production; amino acid; animal feed additive.
 OS Corynebacterium glutamicum.
 PN DE19548222-A1.
 PD 26-JUN-1997.
 PF 22-DEC-1995; 048222.
 PR (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 PI Eggeling L, Sahm H, VriJjc M;
 DR WPI; 97-333867/31.
 DR N-PSDB; T96816.
 PT Increasing microbial production of amino acids, especially lysine -
 PT by improving export carrier activity or corresponding gene
 PT expression, also new export and regulatory genes from
 PT Corynebacterium
 PS Disclosure: Page 9; 16pp; German.
 CC This sequence is the Orf3 protein product, involved in lysine production.
 CC lysG and lysE encode a lysine transport regulatory protein and an export
 CC protein, respectively. Microbial production of amino acids (A) is
 CC improved by increasing the export-carrier activity and/or the export gene
 CC expression in a microorganism that produces (A). The method is
 CC specifically used to increase production of lysine, used as an animal
 CC feed additive. Other (A) are variously useful as pharmaceuticals,
 CC condiments and intermediates for fine chemicals. This method increases
 CC the amount of (A) secreted into the culture medium. Export of (A) has
 CC been found to depend on a single gene.

SQ Sequence 216 AA;
 Query Match 3.4%; Score 90; DB 26; Length 216;
 Best Local Similarity 27.5%; Pred. No. 5.00e+01;
 Matches 22; Conservative 24; Mismatches 28; Indels 6; Gaps 5;
 Db 17 sgkalygissypeltaaef-maeegc--p11lh-qpsylmnrweppddgenn1 72
 58 NSKOTYSGVPIIANDTVGT-FEMAKVICKFSLFVAVHKHSLV-QWDFAGQNPDCLE 115
 Db 73 gsaanglyvafsp1agql 92
 116 HLAASSTGSSDPEQLLEQL 135

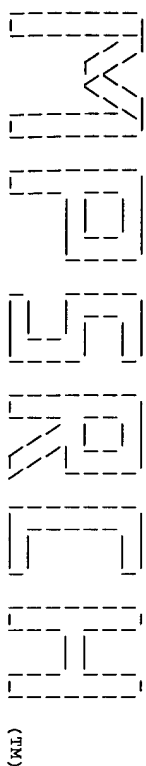
RESULT 14
 ID W34892 standard; Protein; 307 AA.
 AC W34892;
 DT 03-MAR-1998 (first entry)
 DE Novel human phosphorylase kinase gamma subunit.
 KM Phosphorylase kinase gamma subunit; replacement therapy;
 KM glycogen utilisation disorder; muscle glycoemias; hypoglycaemia;
 OS Homo sapiens.
 FH Key
 FH Domain
 FT Location/Qualifiers
 FT 1..300
 FT /note= "putatively contains the catalytic kinase
 FT domain"
 FT 300..385
 FT /note= "putatively contains 2 calmodulin binding
 FT domains required to activate the protein
 FT complex"
 FT Binding_site
 FT 111
 FT /note= "putative substrate binding site"
 FT Binding_site
 FT 154
 FT /note= "putative substrate binding site"
 FT Active_site
 FT 155
 FT /note= "important for catalytic activity"
 FT 168..170
 FT /note= "important for catalytic activity"
 PN US5683910-A.
 PD 04-NOV-1997.
 PF 13-SEP-1996; 713828.
 PR 13-SEP-1996; US-713828.
 PA (INCYTE-) INCYTE PHARM INC.
 PI Bandman O, Goli SK;
 DR WPI; 97-548985/50.
 DR N-PSDB; T93887.
 PT New nucleic acid encoding human phosphorylase kinase gamma subunit -
 PT used for replacement, or gene, therapy of glycogen utilisation
 PT disorders, also for drug screening and diagnosis
 PS Claim 1; Fig 1; 25pp; English.
 CC The present sequence represents a novel human phosphorylase kinase gamma
 CC subunit, characterised as having homology to other phosphorylase kinase
 CC gamma subunits. Recombinant human phosphorylase kinase gamma subunit is useful
 CC for replacement therapy in cases of glycogen utilisation disorders,
 CC e.g. muscle glycoemias, causing reduced growth and kidney damage. Cells
 CC transformed with the DNA encoding the present sequence can also be used
 CC to screen for modulators of human phosphorylase kinase gamma subunit
 CC activity, i.e. potential therapeutic agents. The protein can also be used
 CC similarly in gene therapy, while antisense nucleic acid or ribozymes can
 CC be used to suppress mutant forms of the this gene. The protein can be
 CC used to raise specific antibodies, useful as immunoassay reagents.
 CC Fragments of the DNA encoding the present sequence can be used
 CC diagnostically, e.g. to detect or quantify gene expression, to isolate
 CC genes or related nucleic acid, to detect activation/induction of human
 CC phosphorylase kinase gamma subunit-expressing genes and for mapping the
 CC gene.
 SQ Sequence 307 AA;
 Query Match 3.4%; Score 89; DB 26; Length 307;
 Best Local Similarity 30.6%; Pred. No. 5.86e+01;

Matches 15; Conservative 14; Mismatches 17; Indels 3; Gaps 3;
 Db 37 yetasemllvleyaagglfdgcvanred-afkexygrlmqlllegyh 84
 QY 285 YGMSSEMMMK-KTAGGVAKRASEGKIVEPPKG-DVEHTIDILGIR 331

RESULT 15
 ID W03516 standard; Protein; 359 AA.
 AC W03516.
 DT 11-MAR-1997 (first entry)
 DE Prostaglandin DP receptor.
 KW Prostaglandin DP receptor; human; prostanoid receptor; blood platelet;
 KW smooth muscle; nervous tissue; G protein-coupled receptor; modulator;
 KW mouse; prostaglandin-related disease; therapy.
 OS Homo sapiens.
 PN W09623066-A2.
 PD 01-AUG-1996.
 PE 23-JAN-1996; CA0047.
 PR 26-JAN-1995; US-378682.
 PA (MERI) MERCK FROST CANADA INC.
 PI Abramovitz M, Bole Y, Welters K, Sawyer N, Slipetz DM;
 DR WPI; 96-362690/36.
 DR N-PSDB; T37402.
 PT Human prostaglandin DP receptor and related DNA - used to identify
 PT receptor modulators to treat prostaglandin-related diseases.
 PS Claim 2, Page 41: 49pp. English.
 CC This sequence represents the human prostaglandin DP receptor. The DP
 CC receptor is the least ubiquitous and least abundant of the prostanoid
 CC receptors. The DP receptors are thought to be distributed mainly in
 CC blood platelets, smooth muscle of various tissues, and nervous tissue
 CC (including the central nervous system). However, many of the actions and
 CC distribution of the DP receptor is species dependent. This receptor is a
 CC G protein-coupled receptor which can specifically bind prostaglandin
 CC molecules. The DNA encoding this sequence was isolated using primers
 CC (see T37403 and T37404) based on the N-terminal sequence, and an internal
 CC sequence from purified mouse DP. The human prostaglandin DP receptor can
 CC be used to identify modulators of the receptor. The identified
 CC modulators can then be used to treat prostaglandin-related diseases, and
 CC for modulating the effects of prostaglandins on the DP receptor.
 SQ Sequence 359 AA;

Query Match 3.4%; Score 89; DB 20; Length 359;
 Best Local Similarity 24.7%; Pred. No. 5.86e+01;
 Matches 22; Conservative 26; Mismatches 34; Indels 7; Gaps 7;
 Db 212 latvlgamrnyamhrrl-qrhprsc-trdcapradgreaqspgleelhliall 269
 QY 81 MAKVLCRSLFTAVH-KHYSLVQWQEFAGQNPDCLEHLAASGTSDFEQLEQILE-AI 138
 Db 270 mtvlfmc-slpvlyrayyga-f-kvkkek 296
 QY 139 POKY-ICLDVANGYSEHFVEFVKDKRR 166

Search completed: Mon Jun 15 15:58:12 1998
 Job time : 98 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_p protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Jun 15 15:58:31 1998; MasPar time 16.06 Seconds
Tabular output not generated. 832.618 Million cell updates/sec

Title: >US-08-774-169-1
Description: (1-366) from US08774169.pep
Perfect Score: 2632
Sequence: 1 MTCCPALRFTATPRLSAMP.....RRTFFRVYQVNPISBAC 366

Scoring table: PAM 150
Gap 11

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r56
1:p1r1 2:p1r2 3:p1r3 4:p1r4 5:n1r3d

Statistics: Mean 47.551; Variance 98.999; scale 0.480

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	2052	78.0	345	2	B32902	GMP reductase (EC 1.6	0.00e+00
2	1736	66.0	347	2	H64732	GMP reductase (EC 1.6	0.00e+00
3	550	20.9	488	1	DEBSMP	IMP dehydrogenase (EC	2.81e-78
4	532	20.2	481	2	E64623	inosine 5'-monophosph	7.74e-75
5	529	20.1	326	2	C70015	GMP reductase homolog	2.89e-74
6	529	20.1	493	2	JC4372	IMP dehydrogenase (EC	2.89e-74
7	497	18.9	404	2	E70218	IMP dehydrogenase (EC	3.52e-68
8	491	18.7	496	2	G64501	IMP dehydrogenase (EC	4.83e-67
9	476	18.1	488	1	DEECIP	IMP dehydrogenase (EC	3.31e-64
10	474	18.0	488	2	S23226	IMP dehydrogenase (EC	7.89e-64
11	474	18.0	488	1	H64055	IMP dehydrogenase (EC	7.89e-64
12	466	17.7	484	2	B69056	inosine 5'-monophosph	2.54e-62
13	455	17.3	485	2	JC4998	IMP dehydrogenase (EC	2.98e-60
14	432	16.4	327	2	F64626	GMP reductase - Helic	6.06e-56
15	385	14.6	529	2	S72823	IMP dehydrogenase (EC	3.17e-47
16	364	13.8	503	2	UC4999	IMP dehydrogenase (EC	2.25e-43
17	354	13.4	514	2	J70565	IMP dehydrogenase (EC	1.50e-41
18	354	13.4	514	2	A35566	IMP dehydrogenase (EC	1.50e-41
19	349	13.3	514	2	A31997	IMP dehydrogenase (EC	1.22e-40
20	347	13.2	514	2	B31997	IMP dehydrogenase (EC	2.81e-40
21	345	13.1	537	2	SA1064	IMP dehydrogenase (EC	6.47e-40
22	343	13.0	523	2	S59402	probable IMP dehydrog	1.49e-39
23	343	13.0	524	2	S50890	probable IMP dehydrog	1.49e-39

24	339	12.9	523	2	S48997	probable IMP dehydrog	7.89e-39
25	336	12.8	512	2	A53407	IMP dehydrogenase (EC	2.73e-38
26	328	12.5	537	2	S59508	IMP dehydrogenase (EC	7.60e-37
27	310	11.8	514	2	A38668	IMP dehydrogenase (EC	1.27e-33
28	287	7.7	403	2	S53477	probable IMP dehydrog	4.09e-15
29	202	7.6	387	2	S75050	IMP dehydrogenase cha	5.94e-15
30	163	6.2	157	2	S78078	hypothetical protein	5.52e-09
31	125	4.7	92	5	1AK52	IMP dehydrogenase (EC	1.59e-03
32	104	4.0	174	2	S59994	AMP deaminase (EC 3.5	7.71e-01
33	104	4.0	256	2	S59995	AMP deaminase (EC 3.5	7.71e-01
34	104	4.0	605	2	S59996	AMP deaminase (EC 3.5	7.71e-01
35	104	4.0	760	2	A44313	AMP deaminase (EC 3.5	7.71e-01
36	103	3.9	342	2	S51402	probable membrane pro	1.02e+00
37	103	3.9	508	1	A46008	probable serine-type	1.02e+00
38	102	3.9	511	2	A69369	glutamate synthase (g	1.34e+00
39	102	3.9	622	2	E69006	glutamate synthase (N	1.34e+00
40	101	3.8	168	2	A34952	CS3 fibrillar protein	1.76e+00
41	100	3.8	357	2	E64374	hypothetical protein	2.30e+00
42	98	3.7	168	2	S07904	CS3 pilin precursor -	3.93e+00
43	98	3.7	320	5	1GDH8	glycerate dehydrogena	3.93e+00
44	98	3.7	320	5	1GDH8	glycerate dehydrogena	3.93e+00
45	98	3.7	383	2	E69099	pyruvate oxidoreducta	3.93e+00

ALIGNMENTS

RESULT	ENTRY	1	ALIGNMENT
B32902	#type complete		
GMP reductase (EC 1.6.6.8)	- human		
guanosine monophosphate reductase			
#formal_name Homo sapiens	#common_name man		
20-Dec-1989	#sequence_revision 20-Dec-1989		
10-Sep-1997	#text_change		
B32902; A32902; A55285			
A32902			
#authors	Kanno, H.; Huang, I.Y.; Kan, Y.W.; Yoshida, A.		
#journal	Cell (1989) 58:595-606		
#title	Two structural genes on different chromosomes are required for encoding the major subunit of human red cell glucose-6-phosphate dehydrogenase.		
#cross-references	PubMed:89336791		
#accession	B32902		
#molecule_type	mRNA		
#residues	1-345		
#cross-references	GB:M24470; NID:g182866; PID:g182867; GB:M27958		
#accession	A32902		
#molecule_type	protein		
#residues	43-55		
#note	direct peptide sequencing of a single fragment of a chimeric protein with sequence from glucose-6-phosphate 1-dehydrogenase (A40309) and this protein; the chimeric product was later shown to be an artifact of purification and not a natural component of erythrocytes		
A55285			
Kondoh, T.; Kanno, H.; Chang, L.; Yoshida, A.			
Hum. Genet. (1991) 88:219-224			
Genomic structure and expression of human guanosine monophosphate reductase.			
A55285			
preliminary; translation not shown			
1-233, 'A', '235-255, 'F', '257-345			
##cross-references	GB:S73035		
A55176			
#authors	Henikoff, S.; Smith, J.M.		
#journal	Cell (1989) 58:1021-1022		
#title	The human mRNA that provides the N-terminus of chimeric G6PD encodes GMP reductase.		
A55177			
Yoshida, A.; Kan, Y.W.			
Cell (1990) 62:11-12			

#title	Origin of "fused" glucose-6-phosphate dehydrogenase.
#contents	annotation
GENETICS	
#gene	GDB: GMPR
##cross-references	GDB:127058; OMIM:139265
#map_position	6pter-6qter
KEYWORDS	NADP; oxidoreductase
SUMMARY	#length 345 #molecular-weight 37415 #checksum 4889
Query Match	78.0%; Score 2052; DB 2; Length 345;
Best Local Similarity	79.7%; Pred. No. 0.00e+00;
Matches	275; Conservative 43; Mismatches 27; Indels 0; Gaps 0;
Db	1 MPRDADIKLDFKQVLLRPKRSLKSAEVDLEKFTFRNSKOTYIGIPLIYANDTVGT 60
Qy	19 MPHIDNDVKLDFKQVLLRPKRSLKSEVDLIRSEFSFRNSKOTYIGVPLIYANDTVGT 78
Db	61 FEMAAYNGSHMFAIRKHYSLDMDKFLANNHPECQANVAVSSGSONDELEKSTILEAV 120
Qy	79 FEMAIVLCKESLFRAVHKHYSLVQMOEPAQNDPCLEHLAASGTSSDFOLEQLEI 138
Db	121 PQVFICLDVANGSESEHFVEFKVLRAPFEBHTIMAGNVYTGMAVEELLISGADITKGV 180
Qy	139 PQVKICLDVANGSESEHFVEFKVLRAPFQHTIMAGNVYTGMAVEELLISGADITKGI 198
Db	181 GPGSVCTTRIKTYGVGYQOLSAVIECADSANGLKGIHISDGCCTCPGDVAKAFGADPYM 240
Qy	199 GPGSVCTTRIKTYGVGYQOLSAVIECADSANGLKGIHISDGCSCPDGVAKAFGADPYM 258
Db	241 LGMFSGTEAGVIRNRKILKLFYGNSSDPRAMKHAAGVAYEASGKTYEVPYKGD 300
Qy	259 LGMFSGTEAGVIRNRKILKLFYGNSSDPRAMKHAAGVAYEASGKTYEVPYKGD 318
Db	301 VENTILDLIGLRSTCTYVGAAKLEKESRATFIRYTOQHNVFS 345
Qy	319 VENTILDLIGLRSTCTYVGAAKLEKESRATFIRYTOQHNVFS 363
RESULT	2
ENTRY	H64732 #type complete
TITLE	GMP reductase (EC 1.6.6.8) - Escherichia coli
ORGANISM	#formal_name Escherichia coli
DATE	12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 23-Jan-1998
ACCESSIONS	H64732; S01671; S45182
REFERENCE	A64720
authors	Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal	Science (1997) 277:1453-1462
#title	The complete genome sequence of Escherichia coli K-12.
#cross-references	MUID:9742617
#accession	H64732
#status	nucleic acid sequence not shown; translation not shown
##molecule_type	DNA
##residues	1-347 #label BLAT
##cross-references	GB:AE000119; GB:U00096; NID:gl1786283; PID:gl1786293; UMGF:00104
REFERENCE	##experimental_source strain K-12, substrain MG1655
author	S01671
#journal	Andrews, S.C.; Guest, J.R.
#title	Biochem. J. (1988) 255:35-43
#accession	Nucleotide sequence of the gene encoding the GMP reductase of Escherichia coli K12.
##cross-references	MUID:89061679
#accession	S01671
##molecule_type	DNA
##residues	1-232, 'AR', 236-347 #label AND
##cross-references	EMBL:X07917; NID:g42615; PID:g42616
REFERENCE	S45181
author	Fujita, N.

```
#submission      submitted to the EMBL Data Library, January 1994
#accession       S55182
#molecule_type  DNA
##residues       1-232, 'A', 236-347 #label Fuf
##cross-references EMBL:D6562; NID:g473770; PID:d1006105; PID:g473772

GENETICS
#gene            guac
#map_position    3 min
KEYWORDS
SUMMARY          oxidoreductase
                  length 347 #molecular-weight 37383 #checksum 5330

Query Match      66.0%; Score 1736; DB 2; Length 347;
Best Local Similarity 68.7%; Pred. No. 0.00e+00;
Matches 237; Conservative 53; Mismatches 54; Indels 1; Gaps 1

Db 2 RIEEDKLGFKVLLRPKRSTKSRSVDYLEQGFHKSGOSGSPPIIANNMDIVGPS 61
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy 21 HNDNVKDKDFKVLLRPKRSTKSKSEVDLTISFESFRSKQTYSGPPIIANNMDIVGTFE 80
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 62 MASALASPDILVAHKKHSVEEMQAFINNSSADVLKHVVASTGTSDADPEKTQILDNP 121
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy 81 MAKVKLCFSLFAVHKRHSLVQMOEFAGNP-DCEHLAASSGTSDDPEQLGLEAP 139
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 122 ALNFCIDYANKISEHFQVFVAKAEAMPRTKTCGNVYTGMCEBLITSCADIYKVGIG 181
   :::::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy 140 QKYICLDVANSYSEHFEEFVDVARKREFPQHITMAGNVVTGMVEELITSGADIKRVGIG 199
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 182 PGSVCTTRKTVGVGPOLSAVTEGCADAHAHGGLGSMIVDSGGCTTPGDVAKAFAGSADFWLL 241
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy 200 PGSVCTTRKTKTIVGPOLSAVTEGCADAHAHGLGHITISDGGCCPGDVAAAFAGADFWLL 259
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 242 GMLAGHEBSGRIVEENGKEFMFLFGMSSESAMRRHVGAVEYRAAEKGTYKLPLRGV 301
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy 260 GGMLAGHSHSGGELLERBGKKTKLFYGMSSEWAMKKRYAGVAEYRASGKITVEYFFKDV 319
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 302 ENTADIIJGLASACTYVGSATLKELTKTTFTIRVOEDENRIFFNN 346
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy 320 EHTIRDILGIRSTCTVGAANKELSRRTPIRVTOOVNPIFSE 364
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 3
ENTRY      DEBSMP #type complete
TITLE      IMP dehydrogenase (EC 1.1.1.205) - Bacillus subtilis
ALTERNATE_NAMES
ORGANISM   inositol-monophosphate dehydrogenase guab
           #format_name Bacillus subtilis
DATE       31-Dec-1991 #sequence_revision 09-Aug-1997 #text_change
           12-Dec-1997

ACCESSIONS
REFERENCE   S66039; S12623; D69638
#authors   Ogasawara, N.; Nakai, S.; Yoshikawa, H.
#journal   DNA Res. (1994) 1:1-14
#title     Systematic sequencing of the 180 kilobase region of the
           Bacillus subtilis chromosome containing the replication
           origin.
#accession S66039
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
##residues 1-488 #label OGA
##cross-references EMBL:D26185; NID:g467326; PID:d1005787; PID:g467399
#note      The nucleotide sequence was submitted to the EMBL Data
           Library, December 1993

REFERENCE   S12623
#authors   Kanzaki, N.; Miyagawa, K.
#journal   Nucleic Acids Res. (1990) 18:6710
#title     Nucleotide sequence of the Bacillus subtilis IMP
           dehydrogenase gene.
#cross-references MUID:91067483
#accession S12623
#molecule_type DNA
##residues 1-27, 'H', 29-479, 'VHRNKALPLGLFGSHQKKTGEFYVDECCOGEFFSD'
#label     KAN
#cross-references EMBL:X55669; NID:g939958; PID:g939959
#note      the authors translated the codon CAG for residue 66 as
```

	REFERENCE	A69580	Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Bolton, G.; Aevedo, V.; Bertero, M.G.; Bessieres, P.; Bolin, A.; Borcherl, S.; Boriss, R.; Boussier, L.; Brans, A.; Braun, M.; Bridgell, S.C.; Bron, S.; Brouillet, S.; Bruchi, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Choi, S.K.; Codani, J.U.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Diesterhoef, A.; Ehrlich, S.D.; Emerson, P.T.; Ehtian, K.D.; Erington, J.; Fabret, C.; Ferrari, E.; Fouger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallero, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Goldofty, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.-J.; Haga, K.; Halech, U.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kashara, Y.; Kleer-Blanchard, L.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinols, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, B.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Poh, T.M.; Portetle, D.; Powell, S.; Prescott, A.M.; Prescan, E.; Puic, P.; Punnett, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.; Schlecht, S.; Schaefer, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serron, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takegi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestr, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassartoli, A.; Valit, A.; Wambolt, R.; Wedler, E.; Wedler, H.; Welleneger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yates, K.; Yoshida, K.; Yoshioka, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
#journal	Nature (1997)	380:249-256	
#title	The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.		
#accession	D69638		
#status	nucleic acid sequence not shown; translation not shown		
#molecule_type	DNA		
#residues	1-488 #label KUN		
#experimental_source	strain 168		
GENETICS			
#gene	guab		
CLASSIFICATION	#superfamily IMP dehydrogenase		
KEYWORDS	GMP biosynthesis; NAD: oxidoreductase; purine nucleotide biosynthesis		
FEATURE			
308			
SUMMARY	#active_site Cys #status predicted		
	#length 488 #molecular_weight 52990 #checksum 7257		
Query Match	20.9%; Score 550; DB 1; Length 488;		
Best Local Similarity	38.5%; Pred. No. 2, 81e-78;		
Matches	99; Conservative 62; Mismatches 89; Indels 7; Gaps 5;		
Db	209 EFNNSSKDHLGRLIVGAAGVTG-DTITRVKRKLVEANVDVITDAHGHSQGLNTVTKI	267	
	: ::::: : : : : : : : :		
Qy	105 EFAGQNPDCLLEHLASSGTSDFDEOLEILKA-I-PQVKYICLDVANGSEHFVFVKD	163	
	: ::::: : : : : : : : :		
Db	268 RETPEELINAGNATLEATRALILEAGADVKKGISCTGRVAVAGVQPIRAIDYC	327	
	: ::::: : : : : : : : : :		
Qy	164 KRFPQRTIMGNVTEEMYLELLISADIIKIGDISCVTRTKKTGYGPOLSAVMCC	223	
	: ::::: : : : : : : : : :		
Db	328 ATENARKHGKITIADGGIKTFSGDIITKALAAAGHAVMLGSLLAGTSSESPTETYGRRFXY	387	
	: ::::: : : : : : : : : :		
Qy	224 ADAAHGKLGHILISDGSCPGDVAKAFGACADPYVLGMGLAGHSESGELLIERDDKKYTL	283	
	: ::::: : : : : : : : : :		
Db	388 YRKGSTVAANEKSKDKRYFOEDENKKYPESIEERTIRYKGEVETTYQLVGLRSKGTCG	447	
	: ::::: : : : : : : : : :		
Qy	284 FYGSSSEMANKKYAG--V-AEYR--ASEKTVEPFKGVTEHTINDILLGIIRSTCTYYG	338	
	: ::::: : : : : : : : : :		

RESULT	4	#type complete
ENTRY	E64623	Inosine-5'-monophosphate dehydrogenase - Helicobacter pylori
TITLE	(strain 26695)	
ORGANISM	#formal_name Helicobacter pylori	
DATE	09-Aug-1997	#sequence_revision 09-Aug-1997
ACCESSIONS	E64623	
REFERENCE	E64520	
#authors	Tombl, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.C.; Fleischmann, B.A.; Nelson, K.; Quackenbush, H.P.; Gill, S.; Dougherty, R.D.; Ketchum, K.A.; Klenk, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliak, H.G.; Glodok, A.; McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Mdelman, J.M.; Fujii, C.; Bowman, C.; Wathey, L.; Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M.; Venter, J.C.	
#journal	Nature (1997) 388:539-547	
#title	The complete genome sequence of the gastric pathogen Helicobacter pylori.	
#cross-references	MIMD:9739467	
#accession	E64623	
#status	preliminary; nucleic acid sequence not shown; translation not shown	
##molecule_type	DNA	
##residues	1481	#label TOM
##cross-references	GB:AE000511; NID:g2313957; PID:g2313963	
CLASSIFICATION	#superfamily IMP dehydrogenase	
SUMMARY	#length 481 #molecular_weight 51802 #checksum 462	
Query Match	20.2%; Score 532; DB 2; Length 481;	
Best Local Similarity	38.8%; Pred. No. 7, 74e-75;	
Matches	100; Conservative	65; Mismatches 83; Indels 10; Gaps 7;
Db	204 EYPAKDFGRLVGAIQVGLDRAE-ML-VKAGVDALVDSAGHSANILHLEIK	261
QY	105 EFAGQNDCEHLHASSGSGSPFDELEILAIPOVKYICDANGVSEHFEVEKDVDR	164
Db	262 KSLVVDIV-GNVTKRAITSDLSAGADIKVIGSGSICTRIYAGVCMPOVSAIDMCV	320
QY	165 KRPQHTIMGNVNTGMEVELLSGADITKVGIGSVCTTKRTKGTGVGPOLSAVMCA	224
Db	321 EVASFPDIPYADIGSYSGDVAKALAGASWIMISLTLAGTESPGDMIRQGRQYSY	380
QY	225 DAANGLAGHIISDGGCCPCDVAKARGAGADFMGLGMLAGHSESGELIENDGKRYLF	284
Db	381 RGMGSIAMTKGSSDRYFQEGVASEKLVPEGIEGRVPIYGRKVSMDIFQLVGVRSSMGYQ	440
QY	285 YGMSEAM-K---KT-AGVAEYR-ASEGKTVEVPEFGDVEHTRIDITIGIRSTCTYV	337
Db	441 GARNILELYONAEPVEIT	458
QY	338 GAATKELSKRTTIRYV	355
RESULT	5	
ENTRY	C70015	#type complete
TITLE	GMP reductase homolog yumb - Bacillus subtilis	
ORGANISM	#formal_name Bacillus subtilis	
DATE	05-Dec-1997	#sequence_revision 05-Dec-1997
ACCESSIONS	C70015	
REFERENCE	A69580	

#authors

Kunst, F.; Ogasawara, N.; Moszer, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borcherdt, S.C.; Boris, R.; Bourcier, L.; Brans, A.; Braun, M.; Briganti, S.C.; Bron, R.; Broillet, C.; Bruschi, C.V.; Caldwell, B.; Caputo, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connetton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duestenhoef, A.; Ehrlich, S.D.; Emerson, P.T.; Enlian, K.D.; Erlingsson, U.; Fabre, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goftau, A.; Golightly, E.J.; Grandl, G.; Giuseppe, G.; Guy, B.J.; Haga, A.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holappell, S.; Hosono, S.; Hallö, M.F.; Itaya, M.; Jones, L.; Joris, B.; Kanata, D.; Kasahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koeter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lepidus, A.; Lardinis, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Meunel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Munro, M.; Moestil, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogura, A.; Onodera, B.; Park, S.H.; Patro, V.; Pohlt, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Prescanan, E.; Pujić, P.; Purcell, B.; Rapoport, G.; Ray, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roehle, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeder, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Settori, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Tanaka, T.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tononoli, A.; Tosato, V.; Uchiyama, S.; Vandebout, M.; Vanlier, F.; Vasarotti, A.; Viari, A.; Wanduit, R.; Wedler, E.; Weller, H.; Welleneberger, T.; Winiers, P.; Wispat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshioka, H.F.; Zumbach, E.; Yoshikawa, H.; Danchin, A.

Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
C70015

#accession
#status
#molecule_type DNA
##residues
##experimental_source strain 168

GEMNETS
#gene yumd #length 326 #molecular_weight 35819 #checksum 7996

SUMMARY

Query Match Best Local Similarity 35.3%; Score 529; DB 2; Length 326;
Matches 118; Conservative 79; Mismatches 116; Indels 21; Gaps 17;

D 5 FDYEDIDLPKACVNSRSECD-T-S-V-RIGGHTFK-LPVANMOTITDELAISLAE 59
| : | : | : | : | : | : | : | : | : | : | : | : | :
Q 28 LDENVDLLRPKRSTLKSSVEIDLRSFERNKSOTYSVPPIAAMNMTVGFEAKYLCK 87
| : | : | : | : | : | : | : | : | : | : | : | : | :
D 60 NGFYVMHR-FEPETRIDFIKMNAAGI-FSSISGVXDEEYEVROLAAEBNLPEPVTI 117
| : | : | : | : | : | : | : | : | : | : | : | : | :
Q 88 FSLEFAVKHXYSLWQMGEFA-GONPDCEHLAASGGTSSDFEQLEQLAIPOKVICTCL 146
| : | : | : | : | : | : | : | : | : | : | : | : | :

D 118 DIAHSNAVALEMIOHKLHPDSVFIVANGNGTPAVALEMAADATKTGTSGKYCTIT 177
| : | : | : | : | : | : | : | : | : | : | : | : | :
Q 147 DVANYSHHEFFEFVVDRPKRFPHOIHMGNVTEGMVELLISSADIIRIKYGICSPCTT 206
| : | : | : | : | : | : | : | : | : | : | : | : | :

D 178 KIQTGFSGOMOLAIRCAKA-A-S-K-PITADGIITHRGDIASKIRGAATMWIGSLFA 234
| : | : | : | : | : | : | : | : | : | : | : | : | :
Q 207 RKRTTGVR--QLSAVMECADAHGLKHIIISDGCCSPGVAAAFAGADFWALGGMLTA 264
| : | : | : | : | : | : | : | : | : | : | : | : | :

D 235 GHESPGOTIERDKLTLYKEFGSASEFP--K---G-EK-KNVEKKMHVAHGSIKDPLI 287
| : | : | : | : | : | : | : | : | : | : | : | : | :
Q 265 GHSEGCGELIERDDKRYTLFGMSSEMAMKKYYAGSVAYRAISEGRYTEVPRKGVENTHIR 324
| : | : | : | : | : | : | : | : | : | : | : | : | :

D 288 EMEDDLASSISTYAGTKLNIA-RNDVIYVNKI 320

RESULT ENTRY TITLE IMP dehydrogenase (EC 1.1.1.205) ALTERNAME_NAMES IMP:NAD+ oxidoreductase ORGANISM Streptococcus pyogenes DATE 08-Sep-1997	ACCESSIONS REFERENCE JC4372 #journal Ashbaugh, C.D.; Wessells, M.R. Gene (1995) 165:57-60 Cloning, sequence analysis and expression of the group A streptococcal guab gene encoding inosine monophosphate dehydrogenase.	#accession JC4372 #molecule_type DNA ##residues 1-493 #label ASH ##cross-references GB:026056; NID:g924847; PID:g924848 ##experimental_source TX4 #note The source is designated only as group A Streptococcus.	COMMENT This enzyme is required for putine biosynthesis. GEMNETS #gene guab CLASSIFICATION superfamily IMP dehydrogenase KEYWORDS NAD; oxidoeductase FEATURE 300-313 #region Inosine monophosphate binding #status predicted SUMMARY #length 493 #molecular-weight 52807 #checksum 7848	Query Match Best Local Similarity 40.1%; Score 529; DB 2; Length 493; Matches 99; Conservative 58; Mismatches 80; Indels 10; Gaps 5; DB 225 VAANGVTSDTEREALEFBGADA-IYDTIGHSAAGVLKRIARPPNTLAGN 282 : : : : : : : : : : : : : Q 117 LAASSTGSSEDFOLELOILEAPQKYICLDVANGYSEHFVEDVDRKRRPPOHIIMGN 176 : : : : : : : : : : : : :	DB 283 INABEGARRADYADVNYVVGIPGISICTTRVNAVGVPTWTATDAABAAREKITIIA 342 : : : : : : : : : : : : : Q 177 VTGVEWEELLISGADILVICGPSCVCTRKRKTVPQSJAVECEDAAHAHLKGHITS 236 : : : : : : : : : : : : :	DB 343 DGCIKSYGDIVKALAAGNAVALGSMFAGTDPAPEGTYOGRKFRKTSIAMKKG 402 : : : : : : : : : : : : : Q 237 DGCCSCPDUVAARFAGADPFVLMGLACHSSGSELLERBOKKRLFLYGSSEMAKMK - 295 : : : : : : : : : : : : :	DB 403 SSDRIQGVSUNEANKLVPEIEGRVAYKQAASDIVFMGLGIRSIMGVYAGADIOELHEN 462 : : : : : : : : : : : : : Q 296 ---Y-AGGVAE-YR-ASEGKTIVEYPKODVENHIRTIDLGIRSTCYVGAALKELSRR 348 : : : : : : : : : : : : :	DB 463 AQOVEMS 469 : : : : : : : : : : : : : Q 349 TTFTIRTV 355
---	---	--	---	---	--	--	--	---

RESULT ENTRY TITLE IMP dehydrogenase (EC 1.1.1.205) (guab) - Lyme disease spirochete plasmid B/cp26 FORMAL_NAME Borrelia burgdorferi #common_name Lyme disease spirochete DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 20-Feb-1998 REFERENCES E70218; 140154 AUTHORS Fraser, C.M.; Castagna, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Latifaga, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb, J.E.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; ...	ACCESSIONS REFERENCE E70218; 140154 AUTHORS Fraser, C.M.; Castagna, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Latifaga, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb, J.E.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; ...
--	---

	#journal	J. Utterback, T.; Matthey, L.; McDonald, L.; Attiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, R.; Hatch, B.; Smith, H.O.; Venter, J.C.
	#title	Nature (1997) 390:580-586
	#cross-references	MUID:98065943
	#accession	E70218
	#status	preliminary; nucleic acid sequence not shown;
	##molecule_type	DNA
	##residues	1-404 ##label KLE
	##cross-references	GB:AE000792; TIGR:BBB17
	##experimental_source	Strain B31
	#authors	Margolis, N.; Hogan, D.; Tilly, K.; Rosa, P.
	#journal	J. Bacteriol. (1994) 176:6427-6432
	#title	Plasmid location of Borrelia purine biosynthesis gene homologs
	#cross-references	MUID:95050198
	#accession	I40154
	#status	Preliminary; translated from GB/EMBL/DBJ
	##molecule_type	DNA
	##residues	1-404 ##label RES
	##cross-references	EMBL:U13372; NID:g532791; PID:g532792
	GENETICS	
	#gene	gabA
	CLASSIFICATION	plasmid
	KEYWORDS	#superfamily IMP dehydrogenase oxidoreductase
	SUMMARY	#length 404 #molecular-weight 43767 #checksum 7038
	Query Match	18.9%; Score 497; DB 2; Length 404;
	Best Local Similarity	36.3%; Pred. No. 3.52e-68;
	Matches	95; Conservative 71; Mismatches 86; Indels 10; Gaps 6;
Db	129 EDFENACAD-LNKLIRGAVNSIDITIEREELYKHAVIDLIVDSAHGSTRILEIK	187
Qy	104 QEFAGNDPCLEHLAAASGTSSDFEOLEILAI-PQVKICIDVANGSEHVEFEVKD	162
Db	188 IKTYPNLDLAGNIYRKALDLISVGADCLKGIGSGICTRIYAGVGPQTATCD	247
Qy	163 VRKFPQHTIAGNVGVGEVIELLSGADLIKIGGSVCITTKKTGGVPOLSAWE	222
Db	248 VYEACNNNTNICIIDGGIFRGSDPVAKAIAAGADVMIGMTFAGTRESPEEEIYNKKR	307
Qy	223 CADAHGKGIHIISDGCGSCPDVAKAFACADPVMGLMAGHSSEGELIENDGKRK	282
Db	308 SYVGMSIASAKRSKSKRYOLENNEPKLVPEGIEGNVPYSGLKDIQLQKLGSLSM	367
Qy	283 LFYGMSEMMAK---K-Y-AGVAEYR-AASEKTVFEPKGDVEHTIRIDLIGIRSTC	334
Db	368 YLGAARTSDLKINSFKYKISH	389
Qy	335 TYVGAALKLELSRRRTFLRWTO	356
RESULT	8	
ENTRY	G64501	#type complete
TITLE	IMP dehydrogenase (EC 1.1.1.205) - Methanococcus jannaschii	
ORGANISM	[format_name Methanococcus jannaschii	
DATE	13-Sep-1996	#sequence_revision 13-Sep-1996
DATA	10-Oct-1997	#text_change
ACCESSIONS	G64501	
REFERENCE	A64300	
Authors	Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, A.J.; Goeyayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstein, K.G.; Merrick, J.M.; Glodek, A. Scott, J.L.; Geohagen, N.S.M.; Weidman, J.F.; Fuhmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,	

[illegible]

#title	Nucleotide sequence of the guaB locus encoding IMP dehydrogenase of Escherichia coli K12.
#cross-references	MOID:85215347
#accession	A93555
#molecule	type DNA
#residues	'MOSVTCIMPRXYLLTVLVEILP',1-205,'A',207-488 #label TIE
#cross-references	GB:X02209; NID:g41626; PID:g41627
#experimental	source strain K12
REFERENCE	
#authors	A91329
#journal	Thomas, M.S.; Drabble, W.T.
#title	Gene (1985) 36:45-53
#cross-references	Nucleotide sequence and organisation of the gua promoter region of Escherichia coli.
#accession	MOID:86056959
#molecule	type DNA
#residues	1-96 #label THO
REFERENCE	
#authors	S20017
#journal	Tesfa-Selase, F.; Drabble, W.T.
#title	Mol. Gen. Genet. (1992) 231:256-264
#cross-references	Regulation of the gua operon of Escherichia coli by the DnaA protein.
#accession	MOID:92140365
#molecule	type DNA
#status	not compared with conceptual translation
#residues	1-90 #label TES
COMMENT	The active enzyme, a tetramer of identical chains, catalyzes the conversion of IMP to XMP in the presence of NAD; this reaction is the first unique step to GMP biosynthesis.
GENETICS	
#gene	guaB
#map_position	34 min
CLASSIFICATION	#superfamily IMP dehydrogenase
KEYWORDS	GMP biosynthesis; NAD; oxidoreductase; purine nucleotide biosynthesis
FEATURE	
305	#active_site Cys #status predicted
SUMMARY	#length 488 #molecular-weight 52022 #checksum 7566
Query Match	18.1%; Score 476; DB 1; Length 488;
Best local Similarity	35.9%; Pred. No. 3,31e-64;
Matches	89; Conservative 69; Mismatches 79; Indels 11; Gaps 8;
Db	219 RINGAAGAGAGNERVDALVAGVDV-LT-IDSSHGHSSEVLGRIRTRAKYPDLQIIG 276
QY	116 HLAASGTSGSDEQLEQLLEAIPOVKYICLDVANGXSEHFVFDVKRRFPQHTIMAG 175
Db	277 NVATAGARLAAAGCSAVVGVGSPSICTRTYTGVGVPQITAVADVALEAGTGTPYA 336
QY	176 NVATGEMVELLISGADITKVGIGSPSVCTTRKKTGGVGPQLSAVVECDAAHAGLGH11 235
Db	337 ADGCIFFSGDIAAIAAGASAVVWGSMLAGTESPGIEILYQGRYSKSYRGMSLGAMSK 396
QY	236 SDGCGSCPGVAAAFAGADFPVALGGLAGHSSGELLIRDSKTKKLPFGMS--EMAM 293
Db	397 GSSDRYFQSDNAADKLVPPEGIEGRVAYKGRLEKIIHQOMGLNS-CMGLTGCGTIDELRT 455
QY	294 ---KKY-AGVVAEYR-ASEKGTVEVPFGKDVERTIRDLIGINSTCT-YVGAAKKLKELSR 347
Db	456 KAEFVRIS 463
QY	348 RTTFIRVT 355
ENTRY	S23226
TITLE	IMP dehydrogenase (EC 1.1.1.205) - Acinetobacter calcoaceticus
ORGANISM	formal name Acinetobacter calcoaceticus
DATE	22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 08-Sep-1997
ACCESSIONS	S23226

[illegible]

FEATURE nucleotide biosynthesis

305 #active-site Cys #status predicted

SUMMARY #length 488 #molecular-weight 51980 #checksum 8585

Query Match 18.0%; Score 474; DB 1; Length 488;

Best Local Similarity 36.9%; Pred. No. 7,896-64;

Matches 92; Conservative 68; Mismatches 76; Indels 13; Gaps 10;

Db 220 RVGAAGAGAGNERIDALVAGVDV-LL-IDSSHGSEGLVQVRENRATYPLPIYAG 277

Qy 116 HLAASSGSSDFEOLLEILAPVKYICLDVANGYSEHVEVKKRFPQHTIAG 175

Db 278 NVATAGALADAGASAVKVGISGICTRIYVGVPOITAIADAALAKD-RGIPV 336

Qy 176 NVVGMVEEILISGADIKVIGISGICTRIYVGVPOITAIADAALAKD-RGIPV 336

Db 337 IADGIRPSGDIARALAAAGASCNVGSMFAGTEAPGIEILYQGRKSYKMSGLAMA 396

Qy 235 ISDGSCSPGVAKAFGADFMGLMAGHSESGELIRPDGKRYKLFYGMSS--EMA 292

Db 397 KGSSDRFQSNADKLVEPIEGRIPIYKYLKELIHQOMGLRS-CMGLTGCAITIDELR 455

Qy 293 M---KKY-AGGVAEIR-ASEGKIVEPFGVDEHTIRDLGIRSTCT-YVGAALKLELS 346

Db 456 TRAEFVRIS 464

Qy 347 RRTTFIRVT 355

RESULT 12

ENTRY B69056 #type complete

TITLE inosine-5'-monophosphate dehydrogenase - Methanobacterium

ORGANISM thermoaerophilum (strain Delta H)

DATE #formal_name Methanobacterium thermoaerophilum

05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change

26-Feb-1998

ACCESSION B69056

REFERENCE A69000

AUTHORS Smith, D.R.; Doucette-Stamm, L.A.; Delonghey, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrokowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Moelling, J.; Reeve, J.N.

#journal J. Bacteriol. (1997) 179:7135-7155

#title Complete genome sequence of Methanobacterium thermoaerophilum Delta H: functional analysis and comparative genomics.

#cross-references M01D:96037514

#accession B69056

#status preliminary; nucleic acid sequence not shown; translation not shown

##molecule_type DNA

##residues 1-484 #label MTH

##cross-references GB:AE000666

##experimental_source strain Delta H

GENETICS MTH142

CLASSIFICATION #superfamily IMP dehydrogenase

SUMMARY #length 484 #molecular-weight 52206 #checksum 6744

Query Match 17.7%; Score 466; DB 2; Length 484;

Best Local Similarity 36.1%; Pred. No. 2,546-62;

Matches 97; Conservative 68; Mismatches 88; Indels 16; Gaps 9;

Db 203 ILKRRYRNARSDSGYRVAAAGFPDLER-ARALDE-AGADVLAIDSAGHNKLVKS 260

Qy 100 LVQWOFAGNPDCLHEILAASSGTGSDFEOLLEILAPVKYICLDVANGYSEHVEF 159

Db 261 AGAMKEIEADLIY-GNIATREAAEDLIADVDGLKVGISGSCCTRIIAGVPOLTA 319

Qy 160 VKDVRKRPQHTIAGNVVGMVEEILISGADIKVIGISGICTRIYVGVPOITAIADAALAKD-RGIPV 336

Db 320 IAEVADVAEYGVVINDGIRISGDIARALAAAGASCNVGSMFAGTEAPGIEILYQGRKSYKMSGLAMA 396

Qy 220 VMECADAHGLKHIIDSGCSCGDVAKAFGADFMGLMAGHSESGELIRPDGKRYKLFYGMSS--EMA 292

Db 380 KYKRGMSLGAMTGGISGDIARALAAAGASCNVGSMFAGTEAPGIEILYQGRKSYKMSGLAMA 396

Qy 280 KYKLFYGMSS--EMA 292

Db 440 IGLRASMGCYGAANLSEMERKATVRIT 468

Qy 327 LGIRSTCTYVGAALKLELSRRTTFIRVT 355

RESULT 13

ENTRY JC4998 #type complete

TITLE IMP dehydrogenase (EC 1.1.1.205) - Pyrococcus furiosus

ORGANISM #formal_name Pyrococcus furiosus

DATE 12-Dec-1996 #sequence_revision 21-Jan-1997 #text_change

08-Sep-1997

ACCESSION JC4998

REFERENCE JC4998

AUTHORS Collart, F.R.; Osipuk, J.; Trent, J.; Olsen, G.J.; Huberman, E.

#journal Gene (1996) 174:209-216

#title Cloning, characterization and sequence comparison of the gene coding for IMP dehydrogenase from Pyrococcus furiosus.

#accession JC4998

#status preliminary

##molecule_type DNA

##residues 1-485 #label COL

##cross-references GB:U08814; NID:9595286; PID:9595287

GENETICS

#gene quab

COMPLEX homotetramer

FUNCTION #description provides precursors for DNA and RNA biosynthesis; it catalyzes the conversion of IMP to XMP in the presence of NAD; this reaction is the first unique step to GMP biosynthesis

CLASSIFICATION #superfamily IMP dehydrogenase

KEYWORDS GMP biosynthesis; NAD; oxidoreductase; purine nucleotide biosynthesis

SUMMARY #length 485 #molecular-weight 52900 #checksum 6440

Query Match 17.3%; Score 455; DB 2; Length 485;

Best Local Similarity 36.6%; Pred. No. 2,986-60;

Matches 89; Conservative 58; Mismatches 83; Indels 13; Gaps 8;

Db 222 AAASVPD-LRAIELDRAGVIVDTAHANLKAIRKANKRQKVSADFTV-GNIATP 279

Qy 122 GTGSSDFEOLLEILAPVKYICLDVANGYSEHVEVKKRFPQHTIAGNVVGMVEEILISGADIKVIGISGICTRIYVGVPOITAIADAALAKD-RGIPV 336

Db 280 KAVDDLF--ADAVKVGISGICTRIYVGVPOITAIADAALAKD-RGIPV 336

Qy 181 EMVEELISGADIKVIGISGICTRIYVGVPOITAIADAALAKD-RGIPV 336

Db 338 KYSGDIARALAAAGASCNVGSMFAGTEAPGIEILYQGRKSYKMSGLAMA 396

Qy 241 SCPEVDVAKARAGADFMGLMAGHSESGELIRPDGKRYKLFYGMSS--EMA 292

Db 398 YVGGYKTRKFPVGEVGVVPRYGVSEVLYOLVGLKAGMGVGAIRKELKEGEFV 457

Qy 296 Y-AGGVAEIR-ASEGKIVEPFGVDEHTIRDLGIRSTCTYVGAALKLELSRRTTFI 352

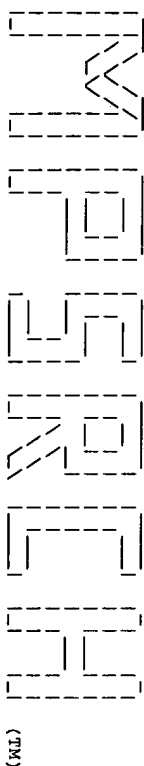
Db 458 IIT 460

Qy 353 RVT 355

ENTRY	14	RESULT	15
TITLE	F64626	#type complete	
ORGANISM	GMP reductase Helicobacter pylori (strain 26655)		
DATE	09-Aug-1997	#formal_name Helicobacter pylori	
ACCESSIONS	10-Oct-1997	#sequence_revision 09-Aug-1997	#text_change
REFERENCE	F64626		
#authors	A64520		
	Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.P.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Goekey, J.D.; Utterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Wathey, L.; Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.M.; Venter, J.C.		
#journal	Nature (1997) 388:539-547		
#title	The complete genome sequence of the gastric pathogen Helicobacter pylori.		
#cross-references	MDID:97394467		
#accession	F64626		
#status	preliminary: nucleic acid sequence not shown; translation not shown		
#molecule_type	DNA		
#residues	1-327	#label TOM	
#cross-references	GB:AE000596; GB:AE000511; MID:g23j3982; PID:g23j3987; TIGR:HP0854		
SUMMARY	#length 327 #molecular-weight 36038 #checksum 7430		
Query Match	16.4%; Score 432; DB 2; Length 327;		
Best Local Similarity	32.3%; Pred. No. 6, 06e-56;		
Matches	108; Conservative 79; Mismatches 126; Indels 21; Gaps 15;		
D	6 FDYEVOLINKNKCVNSRCD-TTYI-L-GKHAQ-MPIYAPANKQINSEIAFLAE 60		
Q	28 LDFEDVLIRKRKRSLKRSRSEVDLTRFSFRNSKQITSGVPIITAAKNDVIGFFEMAKVICK 87		
D	61 NGFYIIMHR-FDGAARIPFYKKMKKR-QWISSISVGKRECLFEVELAKOGIAADPYITI 118		
Q	88 FSLPTAVHKHYSLVQWQEFAGQNPDLCHLAA-SSTGSSDDEQLEQILEALIPQYKICL 146		
D	119 DIANGHSNVIEMIQRTKTHLPETVIAGNVGTPEAVRLENAGADATKVGIGPEKVCIT 178		
Q	147 DVANGYSEHFEVEFKVDKRRFPQHTIMAGNVYTGEEVEELITSGADILKVGIGPEGSVCT 206		
D	179 KIKTGFTGGMOLALRMCAKAR--K-PIADGGIRTHGDIKVSIRGATVMYMGSLFA 235		
Q	207 RKKTGVCYIP-QLSAVWECDAAHGLKHTISGSGSCPGDVAKAFAGADVIMVGSLA 264		
D	236 GHESSGETKIENGIAVKEFGSASEF--OK--G-EK-KNIEGKKIMIQHSGSLKDTLV 288		
Q	265 GHSSGGELEIRROGKKLFLYGMSSSEMAKKYAGVAEIRASEGKTVEYPPFGVDYEHIR 324		
D	289 EMHODLOSSISVAGRDLEAI-RKYDVIYVKNIS 321		
Q	325 DILIGIRSTCTYVGAATKELSRRTTFFIRTOOV 358		
RESULT	15		
ENTRY	S72823	#type complete	
TITLE	IMP dehydrogenase (EC 1.1.1.205) guaB2 - Mycobacterium lepreae		
ALTERNATE_NAMES	B1620.C3_238 protein		
ORGANISM	#formal_name Mycobacterium lepreae		
DATE	19-Mar-1997	#sequence_revision 25-Apr-1997	#text_change
ACCESSIONS	S72823		
REFERENCE	S72580		
#authors	Smith, D.R.; Robison, K.		
#submission	submitted to the EMBL Data Library, November 1993		
#description	Mycobacterium lepreae cosmid B1620.		

	accession	S72823
	##molecule_type	DNA
	##residues	1-529 ##label
	##cross-references	EMBL:U00015; NID:g466531; PID:g466544
	GENETICS	
	gene	guab2
	CLASSIFICATION	superfamily IMP dehydrogenase
	KEYWORDS	GMP biosynthesis; NMD; oxidoreductase; purine nucleotide biosynthesis
	SUMMARY	#length 529 #molecular-weight 54814 #checksum 2745
	Query Match	14.6%; Score 385; DB 2; Length 529;
	Best Local Similarity	35.7%; Pred. No. 3,17e-47;
	Matches	91; Conservative 69; Mismatches 75; Indels 20; Gaps
Db	255	VGAAGVGGDAWBRAMLVDAGDV-LI-VDTAHHRLVLDWYKIKVEIGDRVYIG 3122
Qy	117	LAASGSGSSDFEDELLEIPQYKICIDVANGYSEHVEFEKVDKRRFPQHT-IMAG 1757
Db	313	NVARSAALALVEAGDAWVYVPGSGTCTTRVAVGAGAPQITALE-AVAACGPACVPV 3711
Qy	176	NVATGEMVEEELISGADIIKVGIDPGSVCTTRKKTGVGPQLSVMPCADAPAGLKK-HI 2343
Db	372	IADGGLQYSGDIAKALAGASTYMLGSLAGTAEPPELLIFVNGKOPKSYRCGMSLGAMQ 4313
Qy	235	ISDGGSCPGDVAAAFAGADPFVLLGMLAGHSSGGELLIERDCKKRLKLYGMS--EM- 291
Db	432	GRGDKSTSKDRIYADADLSEDKIVPBGIEGRVDFRQPLSSVILHQLVGLRAANGYTCSP 4911
Qy	292	AM--K--K--Y-AG-GVAEYR-ASEGKTVEVPEFKGDVEHTIIDIIIGIRSTCTYGA 3404
Db	492	TIEVL-QOAPFRIT 505
Qy	341	KLELSRRTTIRVT 355

Search completed: Mon Jun 15 15:59:38 1998
Job time : 67 secs.



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_dp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Jun 15 15:59:54 1998; Maspar time 10.95 Seconds
Tabular output not generated. 838.649 Million cell updates/sec

Title: >US-08-774-169-1
Description: (1-366) from US08774169.pep
Sequence: 1 MTCCLPALRFIATPRLSAMP.....RTTFIRVYQVNPFISENC 366

Scoring table:
Gap 11
PAM 150

Searched: 69111 seqs, 25083644 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot35
1:swiss1

Statistics: Mean 49.471; Variance 81.711; scale 0.605

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	2048	77.8	345	1	GUAC_HUMAN GMP REDUCTASE (EC 1.6. 0.00e+00	
2	1850	70.3	356	1	GUAC_ASCSU GMP REDUCTASE (EC 1.6. 0.00e+00	
3	1709	64.9	346	1	GUAC_ECOLI GMP REDUCTASE (EC 1.6. 0.00e+00	
4	550	20.9	513	1	IMDH_BACSU INOSINE-5'-MONOPHOSPHA 2.61e-98	
5	532	20.2	481	1	IMDH_HELPY INOSINE-5'-MONOPHOSPHA 5.40e-94	
6	529	20.1	493	1	IMDH_STRPY INOSINE-5'-MONOPHOSPHA 2.82e-93	
7	497	18.9	404	1	IMDH_BORBU INOSINE-5'-MONOPHOSPHA 1.21e-85	
8	491	18.7	496	1	IMDH_MERTU INOSINE-5'-MONOPHOSPHA 3.22e-84	
9	476	18.1	488	1	IMDH_ECOLI INOSINE-5'-MONOPHOSPHA 3.44e-80	
10	474	18.0	488	1	IMDH_ACICA INOSINE-5'-MONOPHOSPHA 3.44e-80	
11	474	18.0	488	1	IMDH_HAEIN INOSINE-5'-MONOPHOSPHA 3.44e-80	
12	455	17.3	485	1	IMDH_PYRPU INOSINE-5'-MONOPHOSPHA 1.05e-75	
13	393	14.9	529	1	IMDH_MYCTU INOSINE-5'-MONOPHOSPHA 3.08e-61	
14	385	14.6	529	1	IMDH_MYCLE INOSINE-5'-MONOPHOSPHA 2.16e-59	
15	367	13.8	503	1	IMH3_CANAL PROBABLE INOSINE-5'-MO 2.94e-55	
16	364	13.9	521	1	IMDH_ARATH INOSINE-5'-MONOPHOSPHA 1.43e-54	
17	356	13.5	454	1	IMDH_PNECA INOSINE-5'-MONOPHOSPHA 9.54e-53	
18	354	13.4	514	1	IMD1_HUMAN INOSINE-5'-MONOPHOSPHA 2.72e-52	
19	354	13.3	514	1	IMD2_MOUSE INOSINE-5'-MONOPHOSPHA 2.72e-52	
20	349	13.3	514	1	IMD2_HUMAN INOSINE-5'-MONOPHOSPHA 3.73e-51	
21	347	13.2	514	1	IMD2_MESAU INOSINE-5'-MONOPHOSPHA 1.06e-50	
22	345	13.1	537	1	IMDH_DROME INOSINE-5'-MONOPHOSPHA 3.02e-50	
23	343	13.0	523	1	IMH2_YEAST PROBABLE INOSINE-5'-MO 8.56e-50	

24	343	13.0	524	1	IMH3_YEAST PROBABLE INOSINE-5'-MO 8.56e-50
25	339	12.9	523	1	IMH1_YEAST PROBABLE INOSINE-5'-MO 6.88e-49
26	336	12.8	512	1	IMDH_TRYBG INOSINE-5'-MONOPHOSPHA 3.28e-48
27	323	12.3	514	1	IMD1_MOUSE INOSINE-5'-MONOPHOSPHA 2.76e-45
28	318	12.1	514	1	IMDH_LEIDO INOSINE-5'-MONOPHOSPHA 2.63e-44
29	272	10.3	503	1	IMDH_TRIFP INOSINE-5'-MONOPHOSPHA 5.05e-34
30	202	7.7	403	1	IMH4_YEAST PUTATIVE INOSINE-5'-MO 2.52e-19
31	198	7.5	141	1	YOD1_MYCTU HYPOTHETICAL PROTEIN C 1.60e-18
32	107	4.1	375	1	YOD1_MYCTU HYPOTHETICAL 38.9 KD P 3.3e-02
33	104	4.0	760	1	AMD2_HUMAN AMP DEAMINASE 2 (EC 3. 9.19e-02
34	103	3.9	508	1	YB9_YEAST PUTATIVE SERINE CARBOX 1.28e-01
35	100	3.8	357	1	Y597_MERTU HYPOTHETICAL PROTEIN M 3.43e-01
36	97	3.7	92	1	DMPK_PSEPU PHENOL HYDROXYLASE P0 8.96e-01
37	98	3.7	168	1	FMS3_ECOLI CS3 FIBRINOL SUBUNIT A 6.52e-01
38	98	3.7	321	1	DHG1_HYPMI GLYCERATE DEHYDROGENAS 6.52e-01
39	97	3.7	375	1	Y09J_MYCLE HYPOTHETICAL 38.9 KD P 8.96e-01
40	95	3.6	199	1	IL11_MOUSE INTERLEUKIN-11 PRECURS 1.68e+00
41	95	3.6	205	1	FIXJ_BRAJA TRANSSCRIPTONAL REGULA 1.68e+00
42	95	3.6	289	1	SYN3_HUMAN SYNTAXIN 3. 1.68e+00
43	94	3.6	401	1	AROC_MYCTU CHROMATASE SYNTHASE (E 2.28e+00
44	94	3.6	1435	1	DP3A_MYCTU DNA POLYMERASE III, AL 2.28e+00
45	96	3.6	1533	1	PDM_DROME MATERNAL PDMILIO PROTE 1.23e+00

ALIGNMENTS

RESULT ID	1	STANDARD:	PRT:	345 AA.
AC	P36959;			
DT	01-JUN-1994 (REL. 29, CREATED)			
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)			
DT	01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)			
DE	GMP REDUCTASE (EC 1.6.6.8) (GUANOSINE 5'-MONOPHOSPHATE OXIDOREDUCTASE).			
GN	GMPR.			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; OC			
OC	EUTHERIA; PRIMATES.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 89336791.			
RA	KANNO H., HUANG I.Y., KAN Y.W., YOSHIDA A.;			
RL	CELL 58:595-606(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 92098099.			
RA	KONDOH T., KANNO H., CHANG L., YOSHIDA A.;			
RL	HUM. GENET. 88:219-224(1991).			
RN	[3]			
RP	VARIANTS.			
RX	MEDLINE: 92098100.			
RA	KONDOH T., KANNO H., CHANG L., YOSHIDA A.;			
RL	HUM. GENET. 88:225-227(1991).			
CC	-1- FUNCTION: IT FUNCTIONS IN THE CONVERSION OF NUCLEOBASE, NUCLEOSIDE AND NUCLEOTIDE DERIVATIVES OF G TO A NUCLEOTIDE, AND IN MAINTAINING THE INTRACELLULAR BALANCE OF A AND G NUCLEOTIDES.			
CC	-1- CATALYTIC ACTIVITY: NADPH + GUANOSINE 5'-PHOSPHATE -> NADP(+) +			
CC	-1- SUBUNIT: HOMOTETRAMER.			
CC	-1- SUBUNIT: HOMOTETRAMER.			
CC	-1- POLYMORPHISM: AT LEAST TWO DIFFERENT ALLELES ARE KNOWN.			
CC	-1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC GMP REDUCTASE AND TO IMP DEHYDROGENASE.			
DR	EMBL: L35304; G529232; -			
DR	EMBL: M27941; G529233; -			
DR	EMBL: M24470; G182867; -			
DR	MIM: 139265; -			
KW	PROSITE: PS00487; IMP DH GMP RED: 1.			
OX	OXIDOREDUCTASE: NADP: POLYMORPHISM.			
FT	BINDING 186 186 IMP (POTENTIAL).			
FT	VARIANT 234 234 A -> T (IN AN ALLELE).			
FT	VARIANT 256 256 F -> I (IN AN ALLELE).			
SO	SEQUENCE 345 AA; 37419 MW; 5DB5E0E CRC32;			

```
Db      1 MPDEFBEKDKVLLRPRKRSTLRSAEYDLMEBYERRSKTYGVVPVVASNMDTGT 60
        |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY      19 MPAIDNDVVKLDFEKVLLRPKRSTLSASEVDLTRSFSEFRSKTYGSGVPIIAANMTVTGT 78
        |||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db      61 FEAEVLAKSLEFTHIKHGYDEMKAFVQRVDSNPQIMQSIGSSISSTDPKLRTVC 120
        ||||| |||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY      79 FEAAKVLCFSLEFAVRKHYSLVLMQEFAGQ--NPDCEHLAASSCTGSSDSEFOEQIIL 135
        ||||| |||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
Db      121 DMIPDELICLDVANGXSEYVDFIRRYREOPFTHTIFACNVYTGAEVEELLISGADYVK 180
        ::|||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
QY      136 EALPVKVICLDVANNGXSEHFEVKEVKVRRPQHTITMAGNVYTGAEVEELLISGADIIL 195
        :|||:::|||||:::|||:::|||:::|||:::|||:::|||:::|||:::|||
```

DR	EMBL, L280321; G4753772.	
DR	EMBL, AE000119; G1786293.	-.
DR	EMBL, L28105; G456045.	
DR	PIR, S01671; S01671.	
DR	PIR, S45182; S45182.	
DR	ECODBASE: H037.4; 6TH EDITION.	
DR	ECODBASE: EG10422; GUAC	
DR	PROSITE, PS00487; IMP_DH_GMP_RED; 1.	
KW	Oxidoreductase; NADP.	
FT	BINDING	186 IMP (POTENTIAL).
FT	NP_BIND	216 NADP (POTENTIAL).

FT	CONFLICT	233	234	AR -> GGG (IN REF. 3).
SO	SEQUENCE	346 AA;	37440 MW;	A14B3D9B CRC32:
	Query Match	64.9%;	Score 1709;	DB 1; Length 346;
	Best Local Similarity	68.4%;	Pred. No. 0.00e+00;	
	Matches	236;	Conservative	52; Mismatches 55; Indels 2; Gaps 2
D	2	RIEDDLKGFDFDYLIRPKRSTLKSRSVDLEERQTFKHSQSGMSGVPILANMDVTGTF	61	
	21	HIDNDVKLDPDFDVLIRPKRSTLKSRSVDLEERQTFKHSQSGMSGVPILANMDVTGTF	80	
D	62	MASLASPELITLAAKHHSVEEMOAFINNSADVILKHMVSTGSDADETRTILDLNP	121	
Q	81	MAKYLCKSLTAAKHHSVEEMOAFINNSADVILKHMVSTGSDADETRTILDLNP	139	
D	122	ALNFPCIDVANGYSSEHFVQFAKAREAPRTICAGNVVTEEMCEELLISGADIVKYGIG	181	
Q	140	QVKYICLDVANGYSSEHFVQFAKAREAPRTICAGNVVTEEMCEELLISGADIVKYGIG	199	
D	182	PGSVCTTVKTVGVGPOLSAVIECADAAHGCGMIVSDGCGTTPGVAKAF-ARAPVML	240	
Q	200	PGSVCTTVKTVGVGPOLSAVIECADAAHGCGMIVSDGCGTTPGVAKAF-ARAPVML	259	
D	241	GGMLAGHEESGGRIVTEENGKFMFLFYGSSSPSAAKHHGVAAEYRAAEGRKVKPLPLGPV	300	
Q	260	GGMLAGHEESGGRIVTEENGKFMFLFYGSSSPSAAKHHGVAAEYRAAEGRKVKPLPLGPV	319	
D	301	ENTARDILGGLRSACTVYAGASRLKELTERTTFFIRYQENRIFNN	345	
Q	320	ENTARDILGGLRSACTVYAGASRLKELTERTTFFIRYQENRIFNN	364	
RESULT	4	STANDARD:	513 AA.	
ID	IMDH_BACSV			
AC	P21879;			
DT	01-MAY-1991 (REL. 18, CREATED)			
DT	01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)			
DE	INHOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP			
DE	DEHYDROGENASE) (IMPDH) (IMPD).			
GN	GUAB OR GNAB.			
OC	BACILLUS SUBTILIS.			
OS	PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE; 91067483.			
RA	KANZAKI N., MIYAGAWA K.I.;			
RL	NUCLEIC ACIDS RES. 18:6710-6710(1990).			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-168;			
RC	MEDLINE; 96051385.			
RA	OGASAWARA N., NAKAI S., YOSHIKAWA H.;			
RL	DNA RES. 1:1-14(1994).			
CC	-1- CATALYTIC ACTIVITY. INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =			
CC	XANTHOSINE 5'-PHOSPHATE + NADH.			
CC	-1- PATHWAY: FIRST REACTION UNIOUE TO GMP PROKARYOTIC IMPDH AND TO			
CC	-1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO			
CC	GMP REDUCTASE.			
CC	-1- SIMILARITY: CONTAINS 2 CBS DOMAINS.			
DR	EMBL; X55669; G39959; -.			
DR	EMBL; D26185; G467399; -.			
DR	PIR; S12623; DEBSMP.			
DR	SUBTILIST; B610073; GUAB.			
DR	PROSITE; PS00487; IMP DH GMP RED. 1.			
KW	OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;			
KB	CBS DOMAIN.			
FT	DOMAIN	92	147	CBS 1.
FT	DOMAIN	155	209	CBS 2.
FT	BINDING	308	308	IMP (POTENTIAL).
FT	CONFLICT	28	28	H -> R (IN REF. 2).
FT	CONFLICT	480	513	VHRRKALPGLGSHOKTGVYDECCSGGFFSSD ->
FT				KESNTYIS (IN REF. 2).

SQ	SEQUENCE	513 AA;	55725 MW;	9818609C CRC32;
	Query Match	20.9%;	Score 550;	DB 1; Length 513;
	Best Local Similarity	38.5%;	Pred. No. 2,61e-98;	
	Matches	99;	Conservative	62; Mismatches 89; Indels 7; Gaps 5
Dd	209	EFPNSSKDIHGRLLVGAAGVGTG-DTMTRVKKLVEANDYIVIDTAHGSGVLNTVTKI	267	
Oy	105	EFAAGNPPCLLEHLAASSGTSSDSEFQLIQLLEAI-PQVKYTICLDVANGSYSHFEVFKDV	163	
Dd	268	RTFYELMINTGNVAATAATRALTEAGADVVYKVGISPSITCTTRVAGCYCPQTALTYDC	327	
Oy	164	KRPFQHTMGNNVTGTGMVEELLILSGDIIKVGIGPSCVTTTKTKTGCVGPOLSAVMCC	223	
Dd	328	ATEARKHGKTIADGIFSGSDITKALAGGHAWMLGSLAGTSDEPGETIYYOGRREKV	387	
Oy	224	ADAANGSLGIHSIDCGSCPDAVAKAFACAGDPFMLGSMGLAHSESSELLERCKKYKL	283	
Dd	388	YRGMSVAAAMEKGSKDRTFOEBENKKEVPGEIEGRTPYKGPELTGYLVGLSGMKGYC	447	
Oy	284	FYGMSSEAMAKMYAGG-V-AEYR--ASEGTVVEPFGDVEHTIRDLGIRSTCTYVG	338	
Dd	448	SKDLRLAEQAFTIMT	464	
Oy	339	AALKELSRRTTFIRVT	355	
RESULT	5	ID MH HELPY	STANDARD:	PRT: 481 AA.
AC	P56088:	01-NOV-1997	(REL. 35, CREATED)	
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)		
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)		
DE	INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP			
DE	DEHYDROGENASE) (IMPDH) (IMPD).			
GN	GUA8 OR HP0829.			
OS	HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).			
OC	PROKARYOTA: GRACILICUTES; SCOTOBACTERIA;			
OC	AEROBIC, MOTILE, HELICAL AND/OR VIBRIOID.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-26695.			
RX	MEDLINE: 97394467.			
RA	TOWH J.-F., WHITE O., KERLAUGE A.R., CLAYTON R.A., SUTTON G.G.,			
RA	FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,			
RA	NELSON K., QUACKENBUSH J., ZHOU L., KIRKNES E.F., PETERSON S.,			
RA	LOEFFLER B., RICHARDSON D., DOSBORN R., KHALIK H.G., GLODER A.,			
RA	MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKER E.K.,			
RA	BERG D.E., GOCAVITE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,			
RA	CORTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,			
RA	HAYES W.S., BORODOVSKIY M., KARP P.D., SMITH H.O., FRASER C.M.,			
RA	VENTER J.C.:			
RU	NATURE 388:539-547(1997).			
CC	-I- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(++) + H(2)O -			
CC	XANTHOSINE 5'-PHOSPHATE + NAH.			
CC	-I- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.			
CC	-I- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).			
CC	-I- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO			
CC	GMP REDUCTASE.			
CC	-I- SIMILARITY: CONTAINS 2 CBS DOMAINS.			
DR	PROSITE: PS00487; IMP_DH_GMP_RED; 1.			
DR	TIGR: HP0829; -.			
FT	DOMAIN	89	142	CBS 1.
FT	DOMAIN	150	204	CBS 2.
FT	BINDING	300	300	IMP (POTENTIAL).
SQ	SEQUENCE	481 AA;	51802 MW;	8c13c5596 CRC32;
	Query Match	20.2%;	Score 532;	DB 1; Length 481;
	Best Local Similarity	38.8%;	Pred. No. 5,40e-94;	
	Matches	100;	Conservative	65; Mismatches 83; Indels 10; Gaps 7;
Dd	204	EYPENAKNDGFRLRGALIGVGDLRAD-ML-VKAGVDALVLSAHSHSNIITLTLEIK	261	
	:	:	:	:

```
QY 105 EFAGNPDCLEHLLAASSGTSDFEQLLEALIPQYKICLDVANGXSEHFVEFKDVR 164
Db 262 KSLVVDYIV-GNVTKRTSLISAGADATKVGISGICSTRIVAGVGMQVAIDNCV 320
QY 165 KRFQHTIMGNVYTGEMVELLISGADILKVGISGVCSTRKKTGTGYQLSAVMECA 224
Db 321 EVASKFIPVADGIRSGVAKALAGASSVMIGSLLASTESPDPFMVYOGROKSY 380
QY 225 DAHGLGHITSDGSCSPGVAKAFGAGADFVIMGLAGHSSGGEILIRDKKTKLF 284
Db 381 RMGSGIATMGSSDRYFQEGVASEKLVPEGIEGRVPRKGVSDMIFQVGVSSMGYQ 440
QY 285 YGMSSEAM-K---KY-AGVAEYR-ASEGKTVVEPKGVDEHTIRDLIGINSTCTYV 337
Db 441 GAKNILELYQNAEFVELT 458
QY 338 GAATKELSLRRTTFRVT 355
```

```
RESULT 6
ID IMDH_STRPY STANDARD: PRT: 493 AA.
AC P50099;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUA8.
OS STREPTOCOCCUS PYOGENES.
OC PROKARYOTA; FIRMICUTES; COCCI; STREPTOCOCCACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96084952.
RA ASHBAUGH C.D., WESSELS M.R.;
RL GENE 165:57-60(1995).
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
GMP REDUCTASE.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
DR EMBL: U26056; G924848; -.
DR PROSITE: P500487; IMP_DH_GMP_RED; 1.
KM OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;
KM CBS DOMAIN.
FT BINDING 310 IMP (POTENTIAL).
SQ SEQUENCE 493 AA; 52807 MW; BD2AF453 CRC32;
```

Query Match 20.1%; Score 529; DB 1; Length 493;
Best Local Similarity 40.1%; Pred. No. 2,82e-93;
Matches 99; Conservative 58; Mismatches 80; Indels 10; Gaps 5;

```
Db 225 VAAAVGTSDFEFERAEALFEAGADA--IVIDTAGHSGAGVARKIAEIRAHFPNFTLIGN 282
QY 117 LAASSGSSDFEQLLEALIPQYKICLDVANGXSEHFVEFKDVRPHTIAGN 176
Db 283 IATLEGARALYDAQVDYKVGISGICSTRIVAGVGMQVAIDNCV 342
QY 177 VVTGEMVEELLISGADILKVGISGVCSTRKKTGTGYQLSAVMECAAHGKHTIS 236
Db 343 DGGIKTSGDIYKALAGANAYLGMFAGTDEAPGETIYGRKFKYRGASIAAMKG 402
QY 237 DGGSCSPDVAKAFGAGADFVIMGLAGHSSGGEILIRDKKTKLFYGMSSEAMK- 295
Db 403 SSDRYFQSGVNEAKLVPEGIEGRVAYKGAASDIYFQMLGIRSGMGYVAGDIQELHEN 462
QY 296 ---Y-AGVAEYR-ASEGKTVVEPKGVDEHTIRDLIGINSTCTYVGAATKELSLR 348
Db 463 AQVEMGS 469
QY 349 TTFIRVT 355
```

```
RESULT 7
ID IMDH_BORBU STANDARD: PRT: 404 AA.
AC P49058;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD) (FRAGMENT).
GN GUA8.
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC PLASMID 26 KB CIRCULAR.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; SPIROCHETES; SPIROCHAETALES;
OC SPIROCHAETACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-B31;
RX MEDLINE: 95050198.
RA MARGOLIS N., HOGAN D., TILLY K., ROSA P.;
RL J. BACTERIOL. 176:6427-6432(1994).
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
GMP REDUCTASE.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
DR EMBL: U13372; G532792; -.
DR PROSITE: P500487; IMP_DH_GMP_RED; 1.
KM OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;
KM CBS DOMAIN; PLASMID.
FT BINDING 229 IMP (POTENTIAL).
FT NON TER 404 404
SQ SEQUENCE 404 AA; 43767 MW; 85988E44 CRC32;
```

Query Match 18.9%; Score 497; DB 1; Length 404;
Best Local Similarity 36.3%; Pred. No. 1.21e-85;
Matches 95; Conservative 71; Mismatches 86; Indels 10; Gaps 6;

```
Db 129 EDFNPACKD-LNKLRTGAVSIDIDIEVEEYKRAHVILYDASHGSTRIELIK 187
QY 104 QEFAGNPDCLEHLLAASSGTSDFEQLLEALIPQYKICLDVANGXSEHFVEFKD 162
Db 188 IKTYPMILDIAGNIVTKEALDLISGADCLKVGISGICSTRIVAGVGMQVAIDNCV 247
QY 163 VKRFPHTIAGNVYTGEMVELLISGADILKVGISGVCSTRKKTGTGYQLSAVME 222
Db 248 VYECNNNTNICIADGIRSGDVYKALAGASSVMIGSLFASTKESPESEIYNGKFK 307
QY 223 CADAHGLGHITSDGSCSPGVAKAFGAGADFVIMGLAGHSSGGEILIRDKKTK 282
Db 308 SYVMGSIAMKRSKSRIFOLENNEFKLVPEGIEGRVAYKGAASDIYFQMLGIRSG 367
QY 283 LFYMSSEAMK--K--Y-AGVAEYR-ASEGKTVVEPKGVDEHTIRDLIGINSTC 334
Db 368 GYLGAATISDKINSKFKVISH 389
QY 335 TYVGAATKELSLRRTTFRVTQ 356
```

```
RESULT 8
ID IMDH_METJA STANDARD: PRT: 496 AA.
AC Q59011;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUA8 OR M1616.
OS METHANOCOCCUS JANNASCHII.
OC ARCHAEABACTERIA; EUBRYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96337999.
RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
```


[illegible]

```

Db 397 GSSDRYFQSDNAADKLVEPEGIEGRVAYKGRLEITHQMGGLRS-CMGLTGCCTIDELRT 455
QY 294 --KTY-AGVAYEYR-ASEGKTVEYFPGVDHTIRDLIGIRSTCT-YVGAAKLELSR 347
Db 456 KAEFRIS 463
QY 348 RTTFIRVT 355

RESULT 10
ID IMDH_AICCA STANDARD; PRT; 488 AA.
AC P31002;
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUAB.
OS ACINETOBACTER CALCOCECTICUS.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
CC NEISSERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EB 104;
RA ANDEREGG U., SCHNUCK W.H., ASPERGER O., KLEBER H.P.;
RL SUBMITTED (JUN-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
CC XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC GMP REDUCTASE.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
DR EMBL; X66859; G38720; -.
DR PIR; S23226; S23226.
DR PROSITE; PS00487; IMP_DH_GMP_RED; 1.
KM OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;
KW CBS DOMAIN.
FT DOMAIN 90 142 CBS 1.
FT BINDING 150 205 CBS 2.
FT SEQUENCE 304 304 IMP (POTENTIAL).
SQ SEQUENCE 488 AA; 51530 MW; 1D4AA72F CRC32;

Query Match 18.0%; Score 474; DB 1; Length 488;
Best Local Similarity 38.9%; Pred. No. 3,44e-80;
Matches 98; Conservative 69; Mismatches 68; Indels 17; Gaps 11;

Db 218 RVGAAGTGTADTPSRVLAEGVGV--IVDTAGHSGAGVIERVYKONFPQOVITGG 275
QY 116 HLAASSGSSDFEQLLEAIPOVKYICLDVANGYSEHFVEYKDYKRRFPQHTIMAG 175
Db 276 NIATGDAALLDAGADAVKVGIGPGSICITRTIYAGIGMPOISAI-DSV-AS-ALKDQIP 332
QY 176 NVVGEWEEIILSGADIIKVGIGSVCTIRKKTGCVGPQLSAMVECADAAHGLKHI- 234
Db 333 LIAAGGIFSGDMKAKIAGASTIVGSLAGTEAPGEVEFPGQRYKAKRAGMSLGAM 392
QY 235 -ISGGCCPGDVAFAFGADFWLGLAGHSGSGELLEIRDOCKKXKLLPYGMSSEVAM 293
Db 393 AGRTGSADRYFQSDNAADKLVEPEGIEGRVAYKGRLEITHQMGGLRS-SMGYTSAYVE 452
QY 294 KRYAGVAEY-RAS-EG--K--T--VE--VPEKGVDEHTIRDLIGIRSTCTIYGAANKLK 343
Db 453 DLRONAKFVKIT 464
QY 344 ELSRRTFIRVT 355

RESULT 11
ID IMDH_HAEIN STANDARD; PRT; 488 AA.
AC P44334;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

```

```

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUAB OR H10221.
OS HAEMOPHILUS INFLUENZAE.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
CC PASTEURRELLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULL C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GGCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
RA WEIDMANN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOW E., COTTON M.D.,
RA UETTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RL SCIENCE 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: INOSINE 5'-PHOSPHATE + NAD(+) + H(2)O =
CC XANTHOSINE 5'-PHOSPHATE + NADH.
CC -1- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC GMP REDUCTASE.
DR EMBL; U32708; G1573185; -.
DR PROSITE; PS00487; IMP_DH_GMP_RED; 1.
DR TIGR; H10221; -.
KM OXIDOREDUCTASE; NAD; GMP BIOSYNTHESIS; PURINE BIOSYNTHESIS; REPEAT;
KW CBS DOMAIN.
FT DOMAIN 91 144 CBS 1.
FT BINDING 152 207 CBS 2.
FT SEQUENCE 306 306 IMP (POTENTIAL).
SQ SEQUENCE 488 AA; 51981 MW; 79CFB63E CRC32;

Query Match 18.0%; Score 474; DB 1; Length 488;
Best Local Similarity 36.9%; Pred. No. 3,44e-80;
Matches 92; Conservative 68; Mismatches 76; Indels 13; Gaps 10;

Db 220 RVGAAGAGAGNERIDALVKAGVDY-LI-IDSSHSGSEGLQVRRTTRAKRPLPIYAG 277
QY 116 HLAASSGSSDFEQLLEAIPOVKYICLDVANGYSEHFVEYKDYKRRFPQHTIMAG 175
Db 278 NVATAGGALADAGASAVKVGIGPGSICITRTIYGVGPQITAIADAALKD-RGIPV 336
QY 176 NVVGEWEEIILSGADIIKVGIGPGSVCTIRKKTGCVGPQLSAMVECADAAHGLKHI 234
Db 337 IADGIRFSGDIKAKIAGASCVMVGMFAGTEAPGEIILYOGRAEKSYRGMSLGAMA 396
QY 235 ISDGGCCPGDVAFAFGADFWLGLAGHSGSGELLEIRDOCKKXKLLPYGMS--EMA 292
Db 397 KGSDDRYFQSDNAADKLVEPEGIEGRVAYKGRLEITHQMGGLRS-CMGLTGATIDELR 455
QY 293 M--KKT-AGVAYEYR-ASEGKTVEYFPGVDHTIRDLIGIRSTCT-YVGAAKLELS 346
Db 456 KAEFRIS 464
QY 347 RTTFIRVT 355

RESULT 12
ID IMDH_PYREFU STANDARD; PRT; 485 AA.
AC P42851;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (EC 1.1.1.205) (IMP
DE DEHYDROGENASE) (IMPDH) (IMPD).
GN GUAB.
OS PYROCOCCLUS FURIOSUS.

```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSEARCH protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Jun 15 16:00:50 1998; MasPar time 19.23 Seconds
Tabular output not generated. 801.463 Million cell updates/sec
Title: >US-08-774-169-1
Description: (1-366) from US08774169.pep
Perfect Score: 2632
Sequence: 1 MTCCDPALRFATPRLSAMP.....RRTRFRVYQVNPISFSEAC 366
Scoring table: PAM 150
Gap 11
Searched: 140542 seqs, 42109429 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: sptrembl5
1:sp_fungi 2:sp_human 3:sp_invertebrate 4:sp_mammal
5:sp_mhc 6:sp_organelle 7:sp_phage 8:sp_plant
9:sp_bacteria 10:sp_rodent 11:sp_virus 12:sp_vertebrate
13:sp_unclassified
Statistics: Mean 47.716; Variance 81.544; scale 0.585
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1852	70.4	358	3	016294	F32D1.5 PROTEIN.	0.00e+00	
2	529	20.1	326	9	005269	HYPOTHETICAL 35.8 KD P	7.63e-91	
3	466	17.7	484	9	026245	INOSINE-5'-MONOPHOSPHA	2.61e-76	
4	454	17.2	478	9	032912	INOSINE-5'-MONOPHOSPHA	1.45e-73	
5	432	16.4	327	9	025525	GMP REDUCTASE (GNAC).	1.49e-68	
6	365	13.9	524	1	014344	HYPOTHETICAL 57.0 KD P	1.74e-53	
7	201	7.6	387	9	P73853	IMP DEHYDROGENASE SUBU	9.89e-19	
8	111	4.2	487	9	033941	ERYBL.	1.09e-02	
9	104	4.0	175	2	016686	AMP DEAMINASE (EC 3.5.	1.19e-01	
10	104	4.0	256	2	016687	AMP DEAMINASE (EC 3.5.	1.19e-01	
11	105	4.0	357	9	051741	COFACTOR MODIFYING PRO	8.50e-02	
12	104	4.0	605	2	016688	AMP DEAMINASE (EC 3.5.	1.19e-01	
13	104	4.0	753	2	016729	AMP DEAMINASE ISOFORM	1.19e-01	
14	103	3.9	342	1	006148	CHROMOSOME XII COSMID	1.66e-01	
15	102	3.9	511	9	029309	GLUTAMATE SYNTHASE (GL	2.30e-01	
16	102	3.9	622	9	026208	GLUTAMATE SYNTHASE (NA	2.30e-01	
17	99	3.8	295	3	001390	SYNTAXIN 1 HOMOLOG.	6.10e-01	
18	100	3.8	307	3	024253	SIMILAR 3. CEREVISIAE	4.42e-01	
19	97	3.7	126	3	024125	CYCLOCHROME P450 (FRAGM	1.15e+00	
20	97	3.7	243	9	031677	XYL PROTEIN.	1.15e+00	

RESULT	1	ALIGNMENTS
ID	016294	PRELIMINARY; PRT; 358 AA.
AC	016294	
DT	01-JAN-1998 (TREMBL,REL. 05, CREATED)	
DT	01-JAN-1998 (TREMBL,REL. 05, LAST SEQUENCE UPDATE)	
DT	01-JAN-1998 (TREMBL,REL. 05, LAST ANNOTATION UPDATE)	
DE	F32D1.5 PROTEIN.	
GN	F32D1.5.	
OS	CAENORHABDITIS ELEGANS.	
OC	ERKAROTIA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BRISTOL N2;	
RX	MEDLINE; 94150718.	
RA	WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,	
RA	BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,	
RA	CRAWTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULFON L.,	
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,	
RA	JONES W., KERSHAW J., KIRSTEN J., LAISTER J., LATREILLE P.,	
RA	LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,	
RA	PARSONS J., PERCY C., RIFEEN L., ROOPRA A., SAUNDERS D., SHOWNKEER R.,	
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,	
RA	THIERRY-MIEG J., THOMAS K., VAUDIN M., VANGHAN K., WATERSTON R.,	
RA	WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDAN P.,	
RU	NATURE 368:32-38(1994).	
RL	[2]	
RN	SEQUENCE FROM N.A.	
RP	STRAIN-BRISTOL N2;	
RC	STRAIN-BRISTOL N2;	
RA	BECKER M., BRADSHAW H., KRAMER J.;	
RL	SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.	
DR	EMBL; AF016427; G2291231. -	
DR	PROSITE; PS00487; IMP_DH_GMP_RED; 1.	
SQ	SEQUENCE 358 AA; 36806 MW; 1A112102 CRC32;	
Query Match	70.4%; Score 1852; DB 3; Length 358;	
Best Local Similarity	70.4%; Pred. No. 0.00e+00;	
Matches 243; Conservative	60; Mismatches 41; Indels 1; Gaps 1;	
Db	1 MPRNEPKLDFKDVLLPRKRSTLKSRAVDYELDRVYFRNSKATYTGVPVVASMDTGT 60	
	HYPOHETICAL 72.6 KD P	1.15e+00
	UDP-N-ACETYLGLUCOSAMIN	8.40e-01
	CHROMOSOME XVI COSMID	1.15e+00
	CODED FOR BY C. ELEGAN	8.40e-01
	GLUTAMYL-TRNA SYNTHETA	2.16e+00
	MO2D9.2.	2.94e+00
	GENOMIC RNA.	2.94e+00
	PUTATIVE SUGAR-BINDING	1.58e+00
	DNA-DIRECTED RNA POLYM	2.94e+00
	29 KDA PROTEIN.	7.31e+00
	CYCLOCHROME C1 (FRAGMEN	5.41e+00
	(STRAN 01/85-PN-B1) P	7.31e+00
	RNA-OR SSDNA-BINDING	5.41e+00
	YCR7 PROTEIN.	4.00e+00
	CYCLOCHROME C1 (FRAGMEN	5.41e+00
	2-KETOISOLVALERATE FER	5.41e+00
	3-PHOSPHOGLYCERATE KIN	7.31e+00
	PROTEIN-SERINE/THREONI	5.41e+00
	HYPOTHETICAL 48.6 KD P	5.41e+00
	HYPOTHETICAL 57.0 KD P	5.41e+00
	CYCLOCHROME P450 (FERUL	7.31e+00
	CD98 HEAVY CHAIN.	4.00e+00
	TRANSLATION INITIATION	4.00e+00
	SENSORY KINASE.	4.00e+00
	T22A3.8.	7.31e+00

[illegible]

```

QY 100 LVQWQFAGONPDCLEHLAASSTGGSDFEQLEILAIPOVKYICLDVANGSEHVEF 159
DB 261 AGAKKEIEADLIY-GNATREAEADLIADVDGLKYIGSGMCTRIIAGVCPOLTA 319
QY 160 VKDVRKFFPHQHTIMAGNVYGVEMWELISGADIIKYGIGSGVCTTRKKTGVYPOLSA 219
DB 320 IAEVADVAAEYGVPIADGIRSGDIKAKAIYAGADCVMLGNLLAGTYEAPGVVYVNGR 379
QY 220 VMECADAAHGLKHHISDGGSCPGDVAKAFAGADFVMLGMLAGHSESGELIERDCK 279
DB 380 KYKQYRBMGSLGANTGGTIGACTDRYFQEPBRHMKTKVPEGVGVPIRYGTSEVLFOL 439
QY 280 KYKLEFVEMSSMAKKKAYGVAE-Y-FASEG--K-T-----VE--VPFKGDVEHTIRDI 326
DB 440 IGLRASMGCYCAANTSEMKEKRAIVRT 468
QY 327 LGGIRSTCTYGAAKAKELSRRTFIRYT 355

RESULT 4
ID 032912 PRELIMINARY; PRT: 478 AA.
AC 032912:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE.
GN GUAB.
OS MYCOBACTERIUM LEPRAE.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RA SKELTON J., CHURCHER C.M.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA EIGLMEIER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
RL MOL. MICROBIOL. 7:197-206(1993).
DR EMBL; AL008609; E1168632;
SQ SEQUENCE 478 AA; 50383 MW; B8FC1940 CRC32;

Query Match 17.2%; Score 454; DB 9; Length 478;
Best Local Similarity 42.3%; Pred. No. 1,45e-73;
Matches 104; Conservative 50; Mismatches 79; Indels 13; Gaps 10;

DB 214 RLRIGAAVGINGDVAGRAQSL-AEAGVDLVVDTAAGYOVKTELEIKCVASLNLGVPLVA 272
QY 116 HLAASSGTG--SSDEQLEQLEAIPOVKYICLDVANGSEHVEFVDVRKRFPQHIMA 174
DB 273 GNVYSAGSTRLLNAGATIVYGVPGAMCTTRMTGYGRQFSVAULECASAARKLNRYV 332
QY 175 GNVYTGWEVELLISGADIIKYGIGSGVCTTRKKTGVYPOLSAVMECADAAHGLKHHI 234
DB 333 WADGVHPRDVALALALAGASNVMIGSWFAGTYSPGDLMDRDHDOPYKESYGASRAV 392
QY 235 ISDGGSCPGDVAKAFAGADFVMLGMLAGHSESGELI-ERQCKYKLEFYGSSEKAM 293
DB 393 VARTVADSSDFRARKALFDEGISTSRMGLDPDHGVEDLDIHITSGVYSTCTYGAASLTA 452
QY 294 --KRYAGVAE-YR-A--SEG-KT--VEV-PFKGDVEHTIRDIIGIRSTCTYGAANKLK 343
DB 453 ELHKEA 458
QY 344 ELSRRT 349

RESULT 5
ID 035525 PRELIMINARY; PRT: 327 AA.
AC 035525:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

```

```

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DB GMP REDUCTASE (GUAC).
GN HP0854.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA;
OC AEROBIC; MOTILE, HELICAL AND/OR VIBRIOID.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695;
RA TOMB, WHITE, KERLAVAGE, CLAYTON, SUTTON, FLEISCHMANN, KETCHUM, KLENK,
RA GILL, DOUGHERTY, NELSON, QUACKENBUSH, ZHOU, KIRKNESS, PETERSON, LOFUS,
RA RICHARDSON, DODSON, KHALAK, GLODER, MCKENNEY, FITZGERALD, LEE, ADAMS,
RA HICKEY, BERG, GOCAYNE, UTTERBACK, PETERSON, KELLEY, COTTON, WEIDMAN,
RA FUJII, BOWMAN, MATTHEY, WALLIN, HAYES, BORODOVSKY, KARP, SMITH,
RA FRASER VENTER.;
RL NATURE 388:539-547(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-26695;
RA TOMB, WHITE, KERLAVAGE, CLAYTON, SUTTON, FLEISCHMANN, KETCHUM, KLENK,
RA GILL, DOUGHERTY, NELSON, QUACKENBUSH, ZHOU, KIRKNESS, PETERSON, LOFUS,
RA RICHARDSON, DODSON, KHALAK, GLODER, MCKENNEY, FITZGERALD, LEE, ADAMS,
RA HICKEY, BERG, GOCAYNE, UTTERBACK, PETERSON, KELLEY, COTTON, WEIDMAN,
RA FUJII, BOWMAN, MATTHEY, WALLIN, HAYES, BORODOVSKY, KARP, SMITH,
RA FRASER VENTER.;
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB000596; G2313987;
SQ SEQUENCE 327 AA; 36038 MW; PAD410A4 CRC32;

Query Match 16.4%; Score 432; DB 9; Length 327;
Best Local Similarity 32.3%; Pred. No. 1,49e-68;
Matches 108; Conservative 79; Mismatches 126; Indels 21; Gaps 15;

DB 6 FYEDVQILNKNCTVNSRSCD-TTVI-L--GKHAFK-MPIVANNQTIINESIAEFLAE 60
QY 28 LDFKDYILRKRSRSLTKRSSEVLDTRSFSPRNSKQTSYGVPIIANNNDYQTFEMAKYLCK 87
DB 61 NGFYIYHNR-FDGAARFPFKMKKKR-QWISSYGVKKECELFVELAKOGIAPYIYI 118
QY 88 FSLPTAVHKHYSLVQWEEFAGNPDCLEHLAA--SSGSSDFFQLEILAIPOVKYICL 146
DB 119 DIAHGHSNVIEMIORIKTLPETFLVAGNVGTPEAVRELENGADATKYGIGPGKVCIT 178
QY 147 DVANGYSEHVEFVKDVRKFFPHQHTIMAGNVYGVEMWELISGADIIKYGIGSGVCTT 206
DB 179 KITGFGTGWQLAALRWCAKAR--K-PITADGIRTHGDIYKISIRFGATMYMIGSLFA 235
QY 207 RKRKTGVGP--QLSAVMECADAAHGLKHHISDGGSCPGDVAKAFAGADFVMLGMLA 264
DB 236 GHESSEGTIEENGIAKKEFGSASEF--QK---G-EK-KNIGSKKIYIOHKSLDITLV 288
QY 265 GHESSEGELIERQCKRYKLEFYGSSEKAMKRYAGVAEYRASGKTVEVFPKGDVEHTIR 324
DB 289 EHMDDLISSISYAGRDLEAI-RKVDVVIYKNSI 321
QY 325 DILGIRSTCTYGAANKAKELSRRTFIRYTOY 358

RESULT 6
ID 014344 PRELIMINARY; PRT: 524 AA.
AC 014344:
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 57.0 KD PROTEIN.
GN SPBC2F12.14C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-972H-7.
RA WOOD V., RAJANDREAM M.A., BARRELL B.G., SKELTON J., CHURCHER C.M.;

```